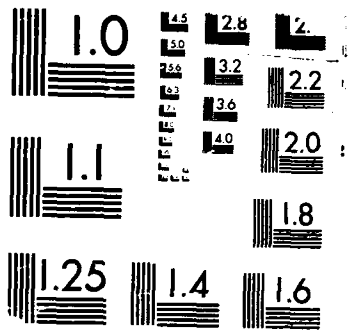
AD-A194 250 ROBUST CONTROL OF MULTIVARIABLE AND LARGE SCALE SYSTEMS 1/2
(U) HONEYWELL SYSTEMS AND RESEARCH CENTER MINNEAPOLIS
MN A PACKARD ET AL. 23 MAR 88 88SRC11 AFOSR-TR-88-0450
UNCLASSIFIED F49620-86-C-0001 F/G 12/3 NL



MICROCOPY RESOLUTION TEST CHART

UREAU STANDARDS-1963-A

A194 250

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ~~0002AA~~ (88SRC11)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR/NM Bolling AFB DC 20332-6448
6a. NAME OF PERFORMING ORGANIZATION Honeywell Inc. 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR-TR- 88-0450
6c. ADDRESS (City, State, and ZIP Code) 3660 Technology Drive Minneapolis, Minnesota 55418	7b. ADDRESS (City, State, and ZIP Code) ~~AFOSR/NM~~ Bldg 410 Bolling AFB DC ~~20332-6448~~
8a. NAME OF FUNDING / SPONSORING ORGANIZATION USAF/AFSC 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-86-C-0001
8c. ADDRESS (City, State, and ZIP Code) Air Force Office of Scientific Research Building 410 Bolling, AFB, DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F · PROJECT NO. 2304 · TASK NO. A1 · WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Robust Control of Multivariable and Large Scale Systems

12. PERSONAL AUTHOR(S)
Andrew Packard and John C. Doyle

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 10/85 TO 3/88	1988 March 23	110

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Robustness, analysis, synthesis, structured singular value, internal stability, linear fractional transformation, H_∞ optimal control.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report gives a fairly complete introduction to the Structured Singular Value (μ) and details some of the latest results. The μ-based methods discussed here have proven to be useful for analyzing the performance and robustness properties of linear feedback systems. This report also describes the recent nonlinear extensions. A description is given for how parametric uncertainty in state space models can be rearranged into the μ framework. The theories for computation of the upper and lower bounds on μ are developed. The relationships between the upper bound and μ for various block structures are explored. The extensions of the μ-based methods to time-varying and nonlinear controllers are outlined.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL James M Crowley, Maj, USAF	22b. TELEPHONE (Include Area Code) (202) 767-5025 · 22c. OFFICE SYMBOL AFOSR/NM

DD Form 1473, JUN 86 — Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

List of Figures

1 Introduction

This report gives a fairly complete introduction to the Structured Singular Value(μ) and details some of the latest results. The μ-based methods discussed here have proven to be useful for analyzing the performance and robustness properties of linear feedback systems. This report also describes the recent nonlinear extensions.

It is assumed that the reader is familiar with the general μ analysis framework. In this context, analysis refers to the process of determining whether a system with a given controller has desired characteristics, whereas synthesis refers to the process of finding a controller that gives desired characteristics, usually expressed in terms of some analysis method. This is the fairly standard usage of these terms in the control community. It should be obvious that the question of analysis must be settled to some degree before a reasonable synthesis problem can be posed. The formal analysis and synthesis techniques discussed are only some of the methods that might make up the overall process of engineering design.

The general framework to be used is illustrated in the diagram in the figure below.

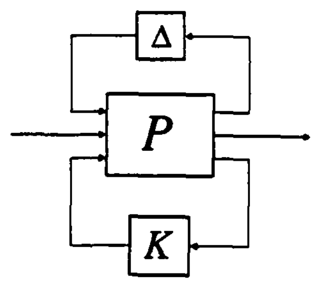


Figure 1.1 General Interconnection

Any linear interconnection of inputs, outputs, commands, perturbations, and a controller can be rearranged to match this diagram. For the purpose of analysis the controller may be thought of as just another system component and the diagram reduces to that below

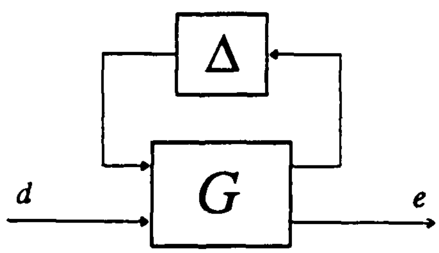


Figure 1.2 Perturbed Disturbance-to-error

The analysis problem involves determining whether the error e remains in a desired set for sets of inputs d and perturbations Δ. Analysis methods differ on the description of these sets and the assumptions on the interconnection structure G. For now, G will be taken to be a linear, time-invariant, lumped system and be represented by a rational transfer function. The convolution kernel associated with G will be denoted as g, so G is a real-rational matrix function of a complex variable and g is a matrix function of time. The interconnection structure G can be partitioned so that the transfer function from d to e can be expressed as the linear fractional transformation

$$e = F_u(G, \Delta)\, d = \left[G_{22} + G_{21}\Delta(I - G_{11}\Delta)^{-1}\, G_{12}\right]\, d.$$

The external input d is an additive signal entering the system and is typically used to model disturbances, commands, and noise. It is generally inadequate in modeling systems for control design to consider uncertainty only in the form of uncertain additive signals. The system model itself typically has uncertainty which can have a significant impact on system performance. This uncertainty is a consequence of unmodeled dynamics and parameter variations and is modeled as the perturbations Δ to the nominal interconnection structure G. Note that the uncertainty modeled as Δ has a very different effect from that of d on the performance of the system. For example, perturbations can cause a nominally stable system to become unstable, which d cannot do.

At the heart of any theory about control are the assumptions made about G, d and Δ, as well as the performance specifications on e. These assumptions determine the analysis methods which can be applied to obtain conclusions about system performance. A desirable objective is to make weak assumptions but still arrive at strong conclusions and the inevitable tradeoff implied by this objective drives the development of new theory. The control theoreticians role may be viewed as one of developing methods that allow the control engineer to make assumptions which seem relatively natural and physically motivated. The ultimate question of the applicability of any mathematical technique to a specific physical problem will always require a "leap of faith" on the part of the engineer and the theoretician can only hope to make this leap smaller.

It is beyond the scope of this report to give a thorough discussion of the relationship between models and the physical systems they represent. Attention will be to the type of models that arise in the μ framework and have proven useful in applications. The particular focus is on techniques that allow very precise analysis of systems which have fairly standard performance requirements and uncertainty models in terms of additive noise and plant perturbations. While the "best" assumptions for engineering purposes will always be a matter of debate, it is clear that for any given set of assumptions it is desirable to have very precise analysis techniques. The ideal would be necessary and sufficient conditions for the satisfaction of a performance specification in the presence of sets of inputs and perturbations. Additionally, the conditions should be computable or should at least yield bounds which give useful estimates of system performance. With such methods, the engineer can focus directly on the relationship between uncertainty assumptions and system performance without worrying about potential gaps caused by inadequate analysis techniques.

The layout is as follows. Section 2 describes how parametric uncertainty in state space models can be rearranged into the μ framework. Section 3 defines μ and its basic properties, along with a few examples. Section 4 is a well known result about an exact expression for μ. Section 5 describes some mathematical preliminaries that are used in subsequent sections concerning the computable upper bound. Section 6 develops theory for the computation of the upper bound, and relates the upper bound to μ. Section 7 explores guaranteed relationships between the upper bound and μ for various block structures. Section 8 is a exposition of linear fractional transformations on structured uncertainties, and how both μ and the upper bound can describe their behavior. Section 9 gives robustness tests for a special class of uncertain difference equations. The extension of the μ-based methods to time-varying and nonlinear controllers is outlined here. Section 10 is a frequency domain/small gain approach to the problem considered in section 9. Section 11 deals with frequency domain μ tests. This material is standard, and is what is usually associated with μ. Section 12 presents counterexamples showing that the upper bound and μ are different. Section 13 describes a power-like algorithm, reminiscent of power algorithms for eigenvalues and singular values, that can be used to get lower bounds for μ. Section 14 is an illustrative example, outlining the various analysis tests and possible conclusions. Finally, Section 15 is the appendix.

2 Parametric Uncertainty in Components

One natural type of uncertainty is unknown coefficients in a state space model. In this section, we will consider a special class of state space models with unknown coefficients, and show how this type of uncertainty can be represented. In particular, we will extract unknown quantities from a parametrically uncertain system so that the perturbations enter the system in a feedback form, or, using the term we will later introduce, in a *linear fractional* way. This type of modeling will form the basic building block for components with **parametric** uncertainty.

After setting up the problem, we will proceed rather informally, manipulating some simple block diagrams to arrive at the special representation of the uncertainty. These types of manipulations are (either explicitly or implicitly) common to the rest of the report particularly section 8. There, while the proofs we give are precise, they tend to hide the key simple idea behind each particular lemma. It is useful to "draw" the block diagrams pertinent to each result, as this makes both the result and proof clearer.

Finally, we reformulate the robustness problem which arises when controlling such uncertain plants into a linear algebra problem, that, eventually, μ will solve. The material of this section is motivated by the discussion in [MorM].

2.1 Problem description

We begin with an explanation of the matrix and block diagram notation that we will use throughout. $\mathbf{C}^{n\times k}$ and $\mathbf{R}^{n\times k}$ are, respectively, all complex and real $n \times k$ matrices. Let $M \in \mathbf{C}^{n\times k}$. As usual, M^T denotes the transpose of M, and M^* denotes the complex conjugate transpose. Suppose u and v are complex vectors, with $u \in \mathbf{C}^k$, $v \in \mathbf{C}^n$, and $v = Mu$. Pictorially, we will draw this relationship as

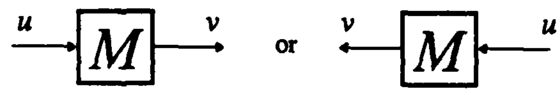


Figure 2.1 Pictorial Notation for Matrix-vector Multiplication

Next, suppose $M \in \mathbf{C}^{(n_1+n_2)\times(k_1+k_2)}$, and we partition in the obvious way as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

with $M_{ij} \in \mathbf{C}^{n_i \times k_j}$. Now if for $i = 1,2$ we have $u_i \in \mathbf{C}^{k_i}$ and $v_i \in \mathbf{C}^{n_i}$, and furthermore

$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = M \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, then we draw this as

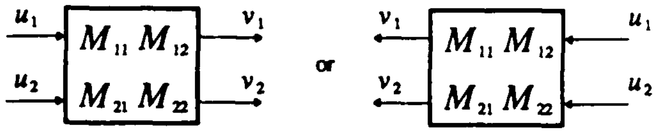


Figure 2.2 Pictorial Notation for Partitioned Matrix-vector Multiplication

When we need the norms of vectors in $\mathbf{C}^n$ or $\mathbf{R}^n$, unless otherwise stated, $\|\cdot\|$ will represent the usual euclidean norm. That is, for $v \in \mathbf{C}^n$, with components $v_i \in \mathbf{C}$, $\|v\|^2 := \sum_{i=1}^{n} |v_i|^2$.

Also, consider a generic finite dimensional, time invariant, linear system, described by

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du. \end{aligned}$$

Note at every instant in time, $\dot{x}, y, u$, and x are related by the simple matrix-vector multiplication

$$\begin{bmatrix} \dot{x} \\ y \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}$$

which in our notation is drawn as

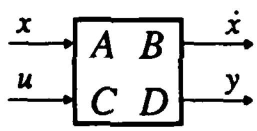


Figure 2.3 Pictorial Notation for Time Invariant, Linear System

Now, onto the problem. Consider a n dimensional, linear system G_δ, parametrized by k uncertain parameters, $\delta_1, \ldots, \delta_k$, and described by the following uncertain equations

$$\begin{aligned} \dot{x} &= \left(A + \sum_{i=1}^{k} \delta_i A_i\right) x + \left(B + \sum_{i=1}^{k} \delta_i B_i\right) u \\ y &= \left(C + \sum_{i=1}^{k} \delta_i C_i\right) x + \left(D + \sum_{i=1}^{k} \delta_i D_i\right) u. \end{aligned} \tag{2.1}$$

Here $A, A_i \in \mathbf{R}^{n \times n}, B, B_i \in \mathbf{R}^{n \times n_u}, C, C_i \in \mathbf{R}^{n_y \times n}$, and $D, D_i \in \mathbf{R}^{n_y \times n_u}$.

The various terms in these state equations are interpreted as follows:

- The nominal system description, given by known matrices A, B, C, and D.
- Parametric uncertainty in the nominal description.

  1. **All** of the uncertainty in the model is contained in the k scalar parameters $\delta_1, \ldots, \delta_k$. Various assumptions on these parameters are possible. For the purposes of this example, we will assume only two things - for each i, $\delta_i \in [-1, 1]$,

and they do not vary with time, they are fixed (though in each instance that the system is operated, the parameters may assume different values, so long as they are in the unit interval).

2. The structural knowledge about the uncertainty is contained in the matrices A_i, B_i, C_i, and D_i. These reflect how the i'th uncertainty, δ_i, affects the state space model. By scaling the entries in these 4 matrices, the relative effect that δ_i has on coefficients is varied. Choosing these matrices is the engineer's job, and is based on her knowledge of the physics that have led to the state space equations.

2.2 Linear fractional transformations

Consider the "perturbed" A matrix (or B or C or D). The jl element of this matrix is of the form $A_{[jl]} + \sum_{i=1}^{k} A_{i_{[jl]}}\delta_i$. Note, that this is **an affine, linear function of the uncertainty.**

Can this model be expressed in the following form?

$$\begin{aligned} \dot{x} &= Ax + Bu + B_2u_2 \\ y &= Cx + Du + D_{12}u_2 \\ y_2 &= C_2x + D_{21}u + D_{22}u_2 \\ u_2 &= \operatorname{diag}[\delta_1 I, \delta_2 I, \ldots, \delta_k I]\, y_2 \end{aligned} \tag{2.2}$$

In other words, can we define some additional inputs, u_2, and outputs, y_2, so that all the uncertainty in the equations (2.1) is represented as a nominal system, G_{nom}, with the unknown parameters entering as the feedback gains that close the loop from the additional outputs to the additional inputs? This is shown in the figure below.

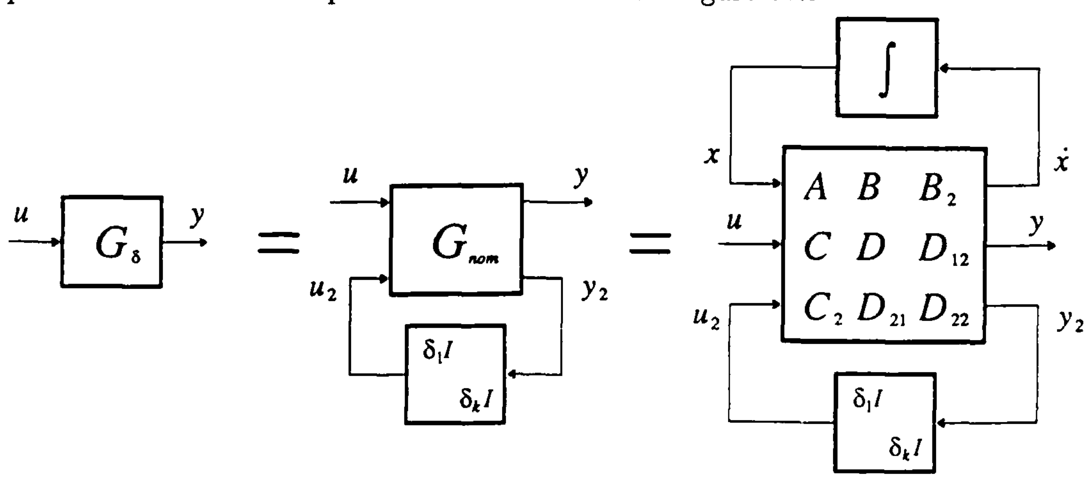


Figure 2.4 Pictorial Notation for Uncertain System

Recall the diagram for the generic linear system. Our problem is then reduced to finding a real matrix M such that for every set of parameters δ_i, the following picture is true,

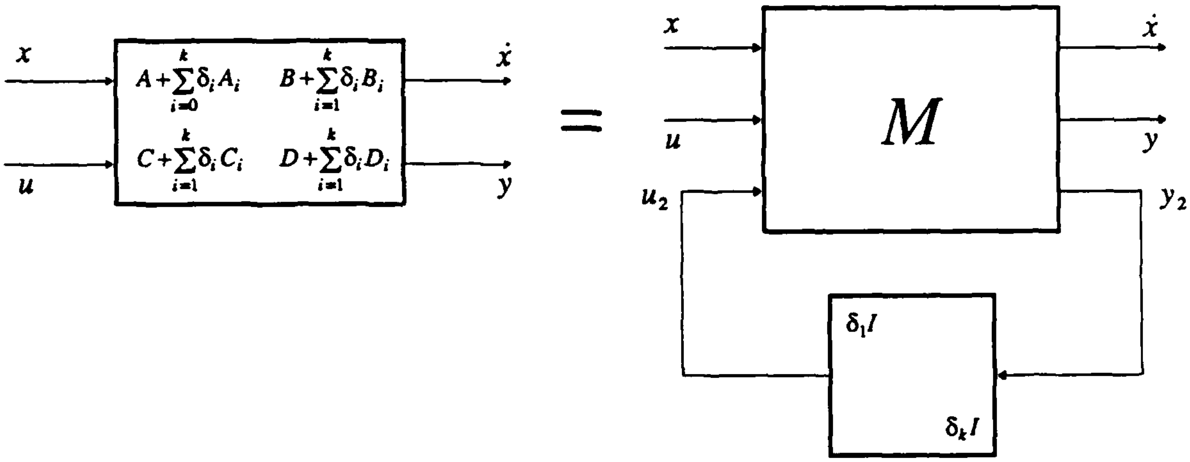


Figure 2.5 Representation of M

In this case, G_{nom} would just be

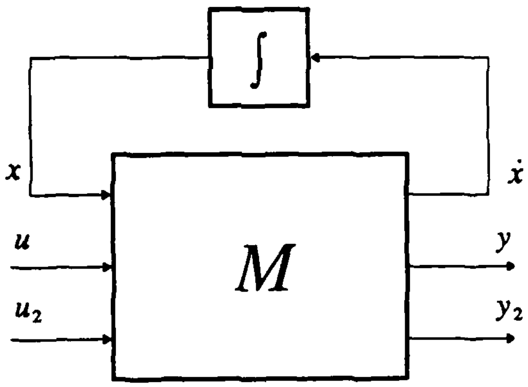


Figure 2.6 Diagram for G_{nom}

Finding such an M is quite easy. Consider a matrix M partitioned in a 2×2 fashion as below left.

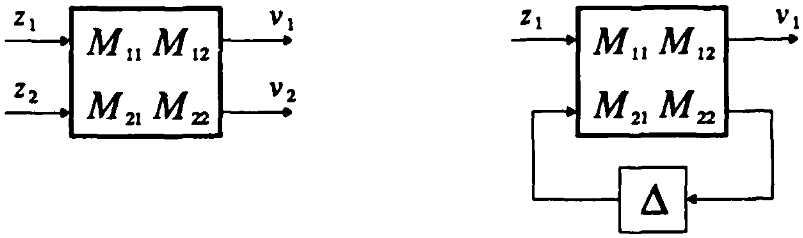


Figure 2.7 Pictorial Notation for a Linear Fractional Transformation

If we close the bottom loop of this with a matrix Δ (as above, right), then the matrix relating z_1 to v_1 is

$$M_{11} + M_{12}\Delta\left(I - M_{22}\Delta\right)^{-1} M_{21}$$

assuming that the inverse exists of course. Since our parameters enter equation (2.1) only affinely, we guess that our M_{22} can be chosen to be zero.

Indeed, for each i, let q_i denote the rank of the matrix

$$P_i := \begin{bmatrix} A_i & B_i \\ C_i & D_i \end{bmatrix} \in \mathbf{R}^{(n+n_y)\times(n+n_u)} \tag{2.3}$$

Then P_i can be written as

$$P_i = \begin{bmatrix} L_i \\ W_i \end{bmatrix} \begin{bmatrix} R_i \\ Z_i \end{bmatrix}^T \tag{2.4}$$

where $L_i \in \mathbf{R}^{n\times q_i}$, $W_i \in \mathbf{R}^{n_y\times q_i}$, $R_i \in \mathbf{R}^{n\times q_i}$, and $Z_i \in \mathbf{R}^{n_u\times q_i}$.

Hence, we have

$$\delta_i P_i = \begin{bmatrix} L_i \\ W_i \end{bmatrix} [\delta_i I_{q_i}] \begin{bmatrix} R_i \\ Z_i \end{bmatrix}^T. \tag{2.5}$$

and therefore "our" $M_{11} + M_{12}\Delta M_{21}$, which is

$$\begin{bmatrix} A + \sum_{i=1}^{k}\delta_i A_i & B + \sum_{i=1}^{k}\delta_i B_i \\ C + \sum_{i=1}^{k}\delta_i C_i & D + \sum_{i=1}^{k}\delta_i D_i \end{bmatrix}$$

in fact looks like

$$\overbrace{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}^{M_{11}} + \overbrace{\begin{bmatrix} L_1 & \cdots & L_k \\ W_1 & \cdots & W_k \end{bmatrix}}^{M_{12}} \overbrace{\begin{bmatrix} \delta_1 I_{q_1} & & \\ & \ddots & \\ & & \delta_k I_{q_k} \end{bmatrix}}^{\Delta} \overbrace{\begin{bmatrix} R_1^T & Z_1^T \\ \vdots & \vdots \\ R_k^T & Z_k^T \end{bmatrix}}^{M_{21}}$$

Therefore, correct definitions for the matrices B_2, C_2, D_{12}, D_{21}, and D_{22} are

$$\begin{aligned} B_2 &= \begin{bmatrix} L_1 & L_2 & \cdots & L_k \end{bmatrix} \\ D_{12} &= \begin{bmatrix} W_1 & W_2 & \cdots & W_k \end{bmatrix} \\ C_2 &= \begin{bmatrix} R_1 & R_2 & \cdots & R_k \end{bmatrix}^T \\ D_{21} &= \begin{bmatrix} Z_1 & Z_2 & \cdots & Z_k \end{bmatrix}^T \end{aligned}$$

and $D_{22} = 0$.

The uncertainty is contained in the block diagonal matrix Δ. We define the "block structure" associated with this system as

$$\mathbf{\Delta} := \{\mathrm{diag}\,[\delta_1 I_{q_1}, \ldots, \delta_k I_{q_k}] : \delta_i \in [-1, 1]\} \tag{2.6}$$

Note that if we had not done the rank reduction (equations 2.3, 2.4, and 2.5), then this structure would, in general, have much larger dimensions.

How would an uncertain parameter enter in a multi-rank way? Consider a system with several different components, each of whose models are affected in a linear fractional way by an something external to the system. For instance, the force/torque producing effectiveness of an airplane's controllable surfaces (rudder, aileron, canard), are affected by ambient dynamic pressure. Suppose that for each surface, the model of its effectiveness has dynamic pressure entering in an affine, linear fashion. Then each surface has an uncertainty associated with pressure. Since these different surfaces affect the airplane in different manners, there is no way to isolate the effect of dynamic pressure as one scalar δ_{bp}. Several of these identical scalars are necessary, and together they form a repeated scalar block.

Remarks: Recall that the uncertain parameters entered both the state equations and ouput equations in an affine, linear fashion. There is a more general model of uncertainty which also leads to the "feedback" representation found in equation (2.2). Each entry in the state space matrices can be a **fraction of affine multilinear combinations of the uncertain parameters**. For example, a particular perturbed entry of one of the matrices may look like

$$\frac{f_{nom} + f_1\delta_1\delta_2 + f_2\delta_3}{1 + h_1\delta_2 + h_2\delta_1\delta_2\delta_3} \tag{2.7}$$

where the f's and h's are known, and represent how the uncertainty affects the matrix entry (our example in this section has all of the h's equal to 0).

These models for uncertainty are called **linear fractional**, and will be explored more in section 8 and 9. Unfortunately, the added generality in (2.7) as compared to (2.1) introduces some difficulties – the nice uncertainty rank reduction procedure (equations 2.3 - 2.5) becomes quite difficult. In fact, it is equivalent to finding minimal realizations of multidimensional (several independent variables) systems. In some simple problems, it is easy to extract the minimal number of uncertainties by inspection. More generally, it is possible that an uncertainty structure much larger (parameters entering many times) than is really necessary is obtained. From a computational viewpoint, this is undesirable.

Also note that any linear connection of several uncertain components (inputs to separate components being linear combinations of outputs of separate components) will have exactly the same form: all of the parametric uncertainty can be isolated in a block diagonal "feedback" around a known, fixed system.

Now, to motivate μ, and the theorems in section 8, suppose we are given an uncertain plant in the form (2.2), and a linear, time invariant, finite dimensional (LTIFD) controller that stabilizes (feeding back y to u) the nominal plant. Under what conditions does it stabilize all of the perturbed plants? First, let the stabilizing controller be governed by $\dot{\zeta} = A_c\zeta + B_c y\ ; u = C_c\zeta$. We have chosen it strictly proper just to simplify some of the equations (all of the robustness questions can be addressed for controllers with D terms). Define the following matrices

$$M_{11} := \begin{bmatrix} A & BC_c \\ B_cC & A_c + B_cDC_c \end{bmatrix} \qquad M_{12} := \begin{bmatrix} B_2 \\ B_cD_{12} \end{bmatrix} \tag{2.8}$$

$$M_{21} := \begin{bmatrix} C_2 & D_{21}C_c \end{bmatrix} \qquad M_{22} := D_{22} \tag{2.9}$$

With $\eta := \begin{bmatrix} x \\ \zeta \end{bmatrix}$, it is straightforward to check that the perturbed closed loop system is

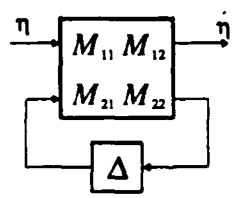


Figure 2.8 Pictorial Notation for Perturbed Closed Loop System

Hence, to guarantee robust stability, we need to verify that for all $\Delta \in \mathbf{\Delta}$ (recall $\mathbf{\Delta}$ is the appropriate uncertainty structure, equation 2.6), the eigenvalues of the matrix

$$M_{11} + M_{12}\Delta\left(I - M_{22}\Delta\right)^{-1} M_{21} \tag{2.10}$$

are in the open left half plane. Alternatively, if the problem had been formulated in discrete time, then the condition would involve making sure the eigenvalues remained inside the unit disc. Actually, this type of test is more directly handled by μ. The μ test (Theorems

9.1 and 9.7) is applied to the whole matrix M, and involves **not only the structure Δ representing the uncertainty**, but an augmented structure which makes sure that the test checks the largest eigenvalue of $M_{11} + M_{12}\Delta\left(I - M_{22}\Delta\right)^{-1} M_{21}$, and not a different quantity, such as the maximum singular value of this perturbed matrix. This is made clearer in section 8 and 9. As we have mentioned though, computation of μ is difficult, and that is the real issue in using any of the results.

2.3 Real vs. complex perturbations

All the theory presented here is appropriate for robustness analysis with **complex** perturbations, and not for **real** perturbations (as in the example in this section). Hence, the typical assumption we will impose on the δ_i in $\boldsymbol{\Delta}$ in (2.6) is actually $\delta_i \in \mathbf{C}, |\delta_i| \leq 1$ for each i. That is, instead of viewing them as fixed unknown real parameters, they are treated as fixed unknown complex parameters. As we will see in section 11, this is also equivalent to treating them as stable, finite dimensional, linear time invariant systems, with $|\delta_i(j\omega)| \leq 1$ for all $\omega \in \mathbf{R}$. Therefore, if a particular problem has uncertainty that is definitely **real** and not dynamical (ie. complex), the methods here will be conservative, since the smallest offending (destabilizing) perturbation will almost always be complex.

It is often very natural to model uncertainty with real perturbations, when, as in this section, the real coefficients of a differential equation model are uncertain. It is important, however, to remember that such parametric variations are in a model, not in the physical system being modeled. Models with real parametric uncertainty are used because, in principle, they allow more accurate representation of some systems. Complex perturbations are typically used to represent uncertainty due to unmodeled dynamics, or to "cover" the variations produced by several real parameters. In the μ framework, complex uncertain blocks also arise for problems of robust performance.

Although computation of μ for complex perturbations is nontrivial and there are important outstanding issues to be resolved, as indicated in this report substantial progress has been made and μ is being applied routinely to large engineering problems. Computation of μ for real perturbations is fundamentally more difficult than for complex perturbations.

The major issues in computing μ, or its equivalent, are the generality of the problem description, the exactness of analysis, and the ease of computation. With existing methods for real perturbations, you get to choose two. A general and, in principle, exact method is a brute force global search using a grid of parameter values (e.g. Horowitz, Ackermann). This inevitably involves an exponential growth in computation as a function of the number of parameters and taking fewer grid points to avoid this gives up exactness. Progress is being made in reducing the computational burden of exact methods ([deGS], [SidG], [SidP]), but nothing suggestive of polynomial-time algorithms is available.

An approach to obtaining exact results with modest computation is to restrict the problem description. The best example is Kharitonov's celebrated result for polynomials with coefficients in intervals. Unfortunately, it is almost impossible to find models with any engineering motivation that fit the allowable problem description. Again, progress is being made in this direction by allowing more general uncertainty descriptions at the expense of more computation.

The approach taken in [FanTD] could be characterized as being very general and computationally attractive, but potentially inexact. Following the methods developed for complex μ, the main idea is to get upper and lower bounds using local search methods which are computationally cheap, but may fail to find global solutions. One then seeks to prove that the local methods yield global solutions, or that the bounds one gets are tight enough to be of value in problems of interest. This strategy has been very successful for complex μ and appears to have promise for the real case as well, although it is clear that the real case is much more challenging.

3 Structured Singular Value

3.1 Definitions

This section is devoted to defining the structured singular value, a matrix function denoted by $\mu(\cdot)$. We consider matrices $M \in \mathbf{C}^{n\times n}$. In the definition of $\mu(M)$, there is an underlying structure $\mathbf{\Delta}$, (a prescribed set of block diagonal matrices) on which everything in the sequel depends. For each problem, this structure is in general different; it depends on the **uncertainty** and **performance objectives** of the problem. Defining the structure involves specifying three things; the type of each block, the total number of blocks, and their dimensions.

There are two types of blocks-*repeated scalar* and *full* blocks. Two nonnegative integers, s and f, represent the number of *repeated scalar* blocks and the number of *full* blocks, respectively. To bookkeep their dimensions, we introduce positive integers $r_1, \ldots, r_s$; $m_1, \ldots, m_f$. The i'th repeated scalar block is $r_i \times r_i$, while the j'th full block is $m_j \times m_j$. With those integers given, we define $\mathbf{\Delta}$ as

$$\mathbf{\Delta} = \left\{\text{diag}\ [\delta_1 I_{r_1}, \ldots, \delta_s I_{r_s}, \Delta_1, \ldots, \Delta_f] : \delta_i \in \mathbf{C}, \Delta_j \in \mathbf{C}^{m_j\times m_j}\right\} \subset \mathbf{C}^{n\times n} \tag{3.1}$$

For consistency among all the dimensions, we must have

$$\sum_{i=1}^{s} r_i + \sum_{j=1}^{f} m_j = n.$$

Often, we will need norm bounded subsets of $\mathbf{\Delta}$, and we introduce the following notation

$$\mathbf{B\Delta} = \{\Delta \in \mathbf{\Delta} : \bar{\sigma}(\Delta) \le 1\} \tag{3.2}$$

Note that in (3.1) we have put all the repeated scalar blocks first. This is just to keep the notation as simple as possible, in fact they can come in any order. In any case, we will see that *every problem can always be set up* (by rearranging rows and columns of M) *so that they appear first*, so we are not losing any generality in this formulation. Also, the *full blocks* do not have to be square, but restricting them as such saves a great deal in terms of notation. This restriction is without loss of generality, since μ for nonsquare blocks can be converted to μ for square blocks by adding rows and/or columns of zeros to M.

Definition 3.1 *For $M \in \mathbf{C}^{n\times n}$, (same dimensions as the elements of $\mathbf{\Delta}$) $\mu_{\mathbf{\Delta}}(M)$ is defined*

$$\mu_{\mathbf{\Delta}}(M) := \frac{1}{\min\limits_{\Delta\in\mathbf{\Delta}} \{\bar{\sigma}(\Delta) : \det(I + M\Delta) = 0\}} \tag{3.3}$$

unless no $\Delta \in \mathbf{\Delta}$ makes $I + M\Delta$ singular, and then $\mu_{\mathbf{\Delta}}(M) = 0$.

An alternative expression follows almost immediately from the definition.

Lemma 3.2 $\mu_{\mathbf{\Delta}}(M) = \max_{\Delta \in \mathbf{B\Delta}} \rho(M\Delta)$

In view of this lemma, continuity of the function $\mu : \mathbf{C}^{n \times n} \to \mathbf{R}$ is apparent. In general, though, the function $\mu : \mathbf{C}^{n \times n} \to \mathbf{R}$ is not a norm, since it doesn't satisfy the triangle inequality. However, for any $\alpha \in \mathbf{C}$, $\mu(\alpha M) = |\alpha| \mu(M)$, so in some sense, it is related to how "big" the matrix is.

We can easily calculate $\mu_{\mathbf{\Delta}}(M)$ when $\mathbf{\Delta}$ is one of two extreme sets.

- If $\mathbf{\Delta} = \{\delta I : \delta \in \mathbf{C}\}$ $(s=1, f=0, r_1=n)$, then $\mu_{\mathbf{\Delta}}(M) = \rho(M)$, the spectral radius of M.

  Proof: The only Δ's in $\mathbf{\Delta}$ which satisfy the $\det(I + M\Delta) = 0$ constraint are negative reciprocals of nonzero eigenvalues of M. The smallest one of these is associated with the largest (in magnitude) eigenvalue, so, $\mu_{\mathbf{\Delta}}(M) = \rho(M)$. $\sharp$

- If $\mathbf{\Delta} = \mathbf{C}^{n \times n}$ $(s=0, f=1, m_1=n)$, then $\mu_{\mathbf{\Delta}}(M) = \bar{\sigma}(M)$

  Proof: If $\bar{\sigma}(\Delta) < \frac{1}{\bar{\sigma}(M)}$, then $\bar{\sigma}(M\Delta) < 1$, so $I + M\Delta$ is nonsingular. Applying equation (3.3) implies $\mu_{\mathbf{\Delta}}(M) \leq \bar{\sigma}(M)$. On the other hand, let u and v be unit vectors satisfying $Mv = \bar{\sigma}(M)u$, and define $\Delta := -\frac{1}{\bar{\sigma}(M)} v u^*$. Then $\bar{\sigma}(\Delta) = \frac{1}{\bar{\sigma}(M)}$ and $I + M\Delta$ is obviously singular. Hence, $\mu_{\mathbf{\Delta}}(M) \geq \bar{\sigma}(M)$. $\sharp$

Obviously, for a general $\mathbf{\Delta}$ as in (3.1) we must have

$$\{\delta I : \delta \in \mathbf{C}\} \subset \mathbf{\Delta} \subset \mathbf{C}^{n \times n}. \tag{3.4}$$

Hence directly from the "minimization" in the definition of μ, and the two simple cases above, we can conclude that

$$\rho(M) \leq \mu_{\mathbf{\Delta}}(M) \leq \bar{\sigma}(M) \tag{3.5}$$

These bounds alone are not sufficient for our purposes, because the gap between ρ and $\bar{\sigma}$ can be arbitrarily large. We refine them by considering transformations on M that **do not affect** $\mu_{\mathbf{\Delta}}(M)$, but **do affect** ρ and $\bar{\sigma}$. To do this, define the following two subsets of $\mathbf{C}^{n \times n}$

$$\mathcal{Q} = \{Q \in \mathbf{\Delta} : Q^* Q = I_n\} \tag{3.6}$$

$$\mathcal{D} = \left\{ \text{diag} \left[D_1, \ldots, D_s, d_1 I_{m_1}, \ldots, d_f I_{m_f} \right] : D_i \in \mathbf{C}^{r_i \times r_i} \text{ is invertible}, d_i \neq 0 \right\} \tag{3.7}$$

Note that for any $\Delta \in \mathbf{\Delta}, Q \in \mathcal{Q}$, and $D \in \mathcal{D}$,

$$Q^* \in \mathcal{Q} \quad Q\Delta \in \mathbf{\Delta} \quad \Delta Q \in \mathbf{\Delta} \quad \bar{\sigma}(Q\Delta) = \bar{\sigma}(\Delta Q) = \bar{\sigma}(\Delta) \tag{3.8}$$

$$D\Delta = \Delta D \tag{3.9}$$

Consequently, we have:

Theorem 3.3 *For all* $Q \in \mathcal{Q}$ *and* $D \in \mathcal{D}$

$$\mu_{\mathbf{\Delta}}(MQ) = \mu_{\mathbf{\Delta}}(QM) = \mu_{\mathbf{\Delta}}(M) = \mu_{\mathbf{\Delta}}\left(DMD^{-1}\right). \tag{3.10}$$

Proof: For all $D \in \mathcal{D}$ and $\Delta \in \mathbf{\Delta}$,

$$\det(I + M\Delta) = \det\left(I + MD^{-1}\Delta D\right) = \det\left(I + DMD^{-1}\Delta\right)$$

since D commutes with Δ. Therefore $\mu_{\mathbf{\Delta}}(M) = \mu_{\mathbf{\Delta}}(DMD^{-1})$. Also, for each $Q \in \mathcal{Q}$, $\det(I + M\Delta) = 0$ if and only if $\det(I + MQQ^*\Delta) = 0$. Since $Q^*\Delta \in \mathbf{\Delta}$ and $\bar{\sigma}(Q^*\Delta) = \bar{\sigma}(\Delta)$, we get $\mu_{\mathbf{\Delta}}(MQ) = \mu_{\mathbf{\Delta}}(M)$ as desired. The argument for QM is the same. ♯

Therefore, the bounds in (3.5) can be tightened to

$$\max_{Q \in \mathcal{Q}} \rho(QM) \leq \mu_{\mathbf{\Delta}}(M) \leq \inf_{D \in \mathcal{D}} \bar{\sigma}(DMD^{-1}) \tag{3.11}$$

An important question is "when are the bounds in (3.11) actually equalities?". This question is a nontrivial one, and a large portion of this report is devoted to answering it. The results we will subsequently show are

- The lower bound, $\max_{Q \in \mathcal{Q}} \rho(QM)$, is always equal to $\mu_{\mathbf{\Delta}}(M)$. Unfortunately, the function $l(Q) := \rho(QM)$ has local maxima which are **not** global, and computing the global maximum of such functions is, in general, impossible.
- In contrast to the local phenomena described above, the function $u(D) := \bar{\sigma}(DMD^{-1})$ does not have any local minima which are not global, so computing $\inf_{D \in \mathcal{D}_e} \bar{\sigma}\left(DMD^{-1}\right)$ is a reasonable task. In general though, $\mu_{\mathbf{\Delta}}(M) < \inf_{D \in \mathcal{D}_e} \bar{\sigma}\left(DMD^{-1}\right)$. For certain block structures $\mathbf{\Delta}$, equality always holds. The general situation is summarized in the table below.

full, $f=$ / δI, $s=$	0	1	2	3	4
0		yes easy	yes Sec. 7.1.1	yes Sec. 7.1.3	no Sec. 7.1.2
1	yes easy	yes Sec. 7.2	no Sec. 12.2	no Sec. 7.1.2	no
2	no Sec. 12.1	no	no	no	no

When is the upper bound, $\inf_{D \in \mathcal{D}_s} \bar{\sigma}\left(DMD^{-1}\right)$, **always** equal to μ ?

The section number in each box indicates where the detailed analysis can be found in this report.

3.2 Simple results & special cases

In this section, we derive simple expressions and bounds for μ in a few special cases. We begin with a class of matrices for which we can derive an easy, explicit expression for μ. This will be done directly from the definition, independent of the upper and lower bounds just described.

Theorem 3.4 *Let n_1, n_2, m_1 and m_2 be positive integers, and consider matrices of the form*

$$\begin{bmatrix} 0 & M_{12} \\ M_{21} & 0 \end{bmatrix} \tag{3.12}$$

where $M_{12} \in \mathbf{C}^{n_1 \times m_2}$, $M_{21} \in \mathbf{C}^{n_2 \times m_1}$ and the zero entries are of the appropriate dimensions. Consider a perturbation set $\boldsymbol{\Delta}$ of the form

$$\boldsymbol{\Delta} = \left\{ \operatorname{diag}\left[\Delta_1, \Delta_2\right] : \Delta_1 \in \mathbf{C}^{m_1 \times n_1}, \Delta_2 \in \mathbf{C}^{m_2 \times n_2} \right\}$$

ie. two full blocks. With respect to this structure,

$$\mu(M) = \sqrt{\bar{\sigma}(M_{12})\,\bar{\sigma}(M_{21})}.$$

Proof: Let M be any matrix as in (3.12), and let $\Delta \in \boldsymbol{\Delta}$. It is straightforward to verify that $\det(I + M\Delta) = \det(I - M_{21}\Delta_1 M_{12}\Delta_2)$. Denote $\sqrt{\bar{\sigma}(M_{12})\,\bar{\sigma}(M_{21})}$ by γ, and suppose that $\Delta \in \boldsymbol{\Delta}$ is chosen with $\bar{\sigma}(\Delta) < \frac{1}{\gamma}$. Then $\bar{\sigma}(M_{21}\Delta_1 M_{12}\Delta_2) < 1$ which means that $I - M_{21}\Delta_1 M_{12}\Delta_2$ is nonsingular, and hence $I + M\Delta$ is nonsingular. This gives a lower bound on the "minimum" part of the definition of μ, namely

$$\min_{\Delta \in \boldsymbol{\Delta}} \{\bar{\sigma}(\Delta) : \det(I + M\Delta) = 0\} \geq \frac{1}{\gamma}$$

Obviously then, from (3.3)

$$\mu(M) \leq \gamma \tag{3.13}$$

Actually (3.13) is an equality; to see this let $u, \tilde{u}, v$ and $\tilde{v}$ be unit vectors of appropriate dimension that satisfy

$$M_{12}v = \bar{\sigma}(M_{12})\,u \quad , \quad M_{21}\tilde{v} = \bar{\sigma}(M_{21})\,\tilde{u}$$

Define the dyads

$$\Delta_1 = \frac{1}{\gamma}\,\tilde{v}u^* \quad , \quad \Delta_2 = \frac{1}{\gamma}\,v\tilde{u}^*.$$

Let $\Delta = \operatorname{diag}[\Delta_1, \Delta_2]$. Obviously $\bar{\sigma}(\Delta) = \frac{1}{\gamma}$, and $(I - M_{21}\Delta_1 M_{12}\Delta_2)\,\tilde{u} = 0$, hence $I + M\Delta$ is singular, and therefore $\mu(M) \geq \gamma$. ♯

The same result was proven in [NetU], using a main result of [Doy]. Here, we used only the definition of μ and simple linear algebra.

The next example gives a easy-to-compute upper bound for rank deficient matrices with arbitrary block structures.

Theorem 3.5 *Suppose $M \in \mathbf{C}^{n\times n}$ has rank r, $r < n$. Then we can write $M = LR^*$, where $L, R \in \mathbf{C}^{n\times r}$. Partition L and R compatibly with the block structure as*

$$L = \begin{bmatrix} L_1 \\ \vdots \\ L_s \\ K_1 \\ \vdots \\ K_f \end{bmatrix} \qquad R = \begin{bmatrix} R_1 \\ \vdots \\ R_s \\ S_1 \\ \vdots \\ S_f \end{bmatrix} \tag{3.14}$$

so that L_i, $R_i \in \mathbf{C}^{r_i\times r}$ and K_i, $S_i \in \mathbf{C}^{m_i\times r}$. Then

$$\mu(M) \leq \sum_{i=1}^{s} \bar{\sigma}(R_i^* L_i) + \sum_{j=1}^{f} \bar{\sigma}(S_j)\,\bar{\sigma}(K_j).$$

Proof: For any $\Delta \in \boldsymbol{\Delta}$

$$\begin{aligned} \det(I + M\Delta) &= \det(I + LR^*\Delta) \\ &= \det(I + R^*\Delta L) \\ &= \det\left(I + \sum_{i=1}^{s} \delta_i R_i^* L_i + \sum_{j=1}^{f} S_j^* \Delta_j K_j\right) \end{aligned} \tag{3.15}$$

If, for some $\beta > 0$, we can show that $\Delta \in \boldsymbol{\Delta}, \bar{\sigma}(\Delta) < \frac{1}{\beta}$ implies that

$$\bar{\sigma}\left(\sum_{i=1}^{s} \delta_i R_i^* L_i + \sum_{j=1}^{f} S_j^* \Delta_j K_j\right) < 1$$

then for all those Δ, $\det(I + M\Delta) \neq 0$ (by (3.15)) and hence $\mu \leq \beta$.

It is easy to find such a β. Suppose that $\Delta \in \mathbf{\Delta}$ and

$$\bar{\sigma}(\Delta) < \frac{1}{\sum_{i=1}^{s} \bar{\sigma}(R_i^* L_i) + \sum_{j=1}^{f} \bar{\sigma}(S_j)\,\bar{\sigma}(K_j)}$$

Then

$$\bar{\sigma}\left(\sum_{i=1}^{s} \delta_i R_i^* L_i + \sum_{j=1}^{f} S_j^* \Delta_j K_j\right) \leq \sum_{i=1}^{s} |\delta_i| \bar{\sigma}(R_i^* L_i) + \sum_{j=1}^{f} \bar{\sigma}(\Delta_j)\,\bar{\sigma}(S_j)\,\bar{\sigma}(K_j) < 1$$

Therefore $\mu(M) \leq \sum_{i=1}^{s} \bar{\sigma}(R_i^* L_i) + \sum_{j=1}^{f} \bar{\sigma}(S_j)\,\bar{\sigma}(K_j)$. ♯

Theorem 3.6 *Let $M \in \mathbf{C}^{n\times n}$ be given, and suppose that M has rank equal to 1. Write $M = LR^*$, and partition L and R compatibly with the block structure as*

$$L = \begin{bmatrix} L_1 \\ \vdots \\ L_s \\ K_1 \\ \vdots \\ K_f \end{bmatrix} \qquad R = \begin{bmatrix} R_1 \\ \vdots \\ R_s \\ S_1 \\ \vdots \\ S_f \end{bmatrix} \tag{3.16}$$

so that $L_i,\ R_i \in \mathbf{C}^{r_i \times 1}$ and $K_i,\ S_i \in \mathbf{C}^{m_i \times 1}$. Then

$$\mu(M) = \sum_{i=1}^{s} |R_i^* L_i| + \sum_{j=1}^{f} \|S_j\|\,\|K_j\|. \tag{3.17}$$

Proof: For notational simplicity, let $\gamma := \sum_{i=1}^{s} |R_i^* L_i| + \sum_{j=1}^{f} \|S_j\|\,\|K_j\|$. Obviously from theorem 3.5 we already have $\mu \leq \gamma$. With M a dyad, we will actually show that it is an equality. For each $i \leq s$, choose $q_i \in \mathbf{C}, |q_i| = 1$, so that $q_i R_i^* L_i$ is a real, nonpositive number. Similarly, for each $j \leq f$, choose a unitary matrix Q_j so that $S_j^* Q_j K_j = -\|S_j\|\,\|K_j\|$. These two steps can always be done. Suppose that $\gamma \neq 0$. Then define

$$\Delta := \frac{1}{\gamma} \operatorname{diag}[q_1 I_{r_1}, \ldots, q_s I_{r_s}, Q_1, \ldots, Q_f] \in \mathbf{\Delta} \tag{3.18}$$

By construction, $\bar{\sigma}(\Delta) = \frac{1}{\gamma}$, and $I + M\Delta$ is singular, therefore $\mu_{\mathbf{\Delta}}(M) \geq \gamma$, so using theorem 3.5, we get the equality as claimed. ♯

4 Proof that lower bound achieves μ

Recall the two bounds we derived in section 3.1.

$$\max_{Q\in\mathcal{Q}} \rho(QM) \leq \mu(M) \leq \inf_{D\in\mathcal{D}} \bar{\sigma}\left(DMD^{-1}\right)$$

A main result of [Doy] is that for any block structure $\boldsymbol{\Delta}$ as defined by (3.1), the left hand side of the bound above is actually an equality:

Theorem 4.1 *Let $\boldsymbol{\Delta}$ be a given block structure, and let the set $\mathcal{Q}$ be defined by (3.6). Then for every matrix M of appropriate dimensions,*

$$\mu(M) = \max_{Q\in\mathcal{Q}} \rho(QM). \tag{4.1}$$

We begin by stating a well known result from complex analysis called Rouche's theorem [Rud].

Theorem 4.2 *Let Γ be a simple closed contour in the complex plane, $\mathbf{C}$. Let f and g be functions which are analytic inside and on Γ. If $|g'z)| < |f(z)|$ on Γ, then f and $f+g$ have the same number of zeros inside Γ.*

This is used in proving the next lemma, which is the well known result stating that the roots of a polynomial are continuous functions of the coefficients of the polynomial.

Lemma 4.3 *Let $f(z) = \sum_{i=0}^{n} a_i z^i$ be an n'th order polynomial, $a_n \neq 0$. Let $\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_n$ be the zeros of f. For any $\epsilon > 0$ and any integer $m > 0$, there exists a $\delta_{m,\epsilon} > 0$ such that if $g(z)$, defined by*

$$g(z) = \sum_{i=0}^{m} b_i z^i$$

has coefficients $b_i \in \mathbf{C}$ which satisfy $|b_i| < \delta_{m,\epsilon}$, then there are n zeros of $f+g$, labeled $\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_n$ that satisfy $|\bar{z}_i - \tilde{z}_i| < \epsilon$.

Hence the zeros of f depend continuously on the coefficients of the polynomial (even leading coefficients which are zero).

Next, we shift our attention to polynomials in several dimensions, that is, polynomials taking $\mathbf{C}^k \rightarrow \mathbf{C}$. If $z \in \mathbf{C}^k$, we let $\|z\|_\infty := \max_{i\leq k} |z_i|$. For $p: \mathbf{C}^k \rightarrow \mathbf{C}$, a polynomial, define β_p as

$$\beta_p = \min\{\|z\|_\infty : p(z) = 0\} \tag{4.2}$$

β_p is the norm of the "smallest" zero of the polynomial. The next lemma is from [Doy].

Lemma 4.4 *Let p be a polynomial from $\mathbf{C}^k \to \mathbf{C}$. Define β_p via (4.2). Then there exists a $z \in \mathbf{C}^k$ such that $|z_i| = \beta_p$ for each i, and $p(z) = 0$.*

Proof: Let $\hat{z}$ be a minimizing solution, so $p(\hat{z}) = 0$ and $\|\hat{z}\|_\infty = \beta_p$. If $|\hat{z}_i| = \beta_p$ for all i, then we are done, so assume that $|\hat{z}_r| < \beta_p$. Now we can (always) write

$$p(z) = \sum_{i=0}^{n} p_i\left(z_1, \ldots, z_{r-1}, z_{r+1}, \ldots, z_k\right) z_r^i \tag{4.3}$$

where the p_i are polynomials in all the variables except z_r.

For notational purpose, we denote $\hat{p}_i$ as the polynomial p_i evaluated at $\hat{z}$ (of course, it doesn't depend on $\hat{z}_r$), that is

$$\hat{p}_i := p_i\left(\hat{z}_1, \ldots, \hat{z}_{r-1}, \hat{z}_{r+1}, \ldots, \hat{z}_k\right)$$

and we let $\mathbf{L}$ denote the set of integers $\{1, 2, \ldots, r-1, r+1, \ldots, k\}$.

There are three situations we need to consider:

1. Suppose that for every i, $\hat{p}_i = 0$. Then, **regardless of the value of** $\hat{z}_r, p(\hat{z}) = 0$. In particular, the magnitude of $\hat{z}_r$ may be adjusted to be β_p and $\hat{z}$ will still be a root of p.

2. Suppose that $\hat{p}_0 \neq 0$, but $\hat{p}_i = 0$ for $i \geq 1$. Quick checking reveals that **this is not possible**, since then $p(\hat{z}) \neq 0$ as we need.

3. Suppose that for some $i \geq 1$, $\hat{p}_i \neq 0$. Then $\hat{z}_r$ is a zero of the nontrivial polynomial $q(z_r) = \sum_{i=0}^{n} \hat{p}_i z_r^i$. Let $\epsilon > 0$ with $|\hat{z}_r| + \epsilon < \beta_p$. By the lemma, we can find a $\delta > 0$ such that if $|\tilde{q}_i - \hat{p}_i| < \delta$ for each i, then the polynomial $\tilde{q}(z_r) := \sum_{i=0}^{n} \tilde{q}_i z_r^i$ would have a zero $\bar{z}_r$ satisfying $|\bar{z}_r - \hat{z}_r| < \epsilon$. Since the p_i are continuous functions of their $k-1$ arguments, we can find a $\bar{\delta} > 0$ such that if $|\zeta_i - \hat{z}_i| < \bar{\delta}$ for all $i \in \mathbf{L}$, then there is a $\bar{z}_r$ with $|\bar{z}_r - \hat{z}_r| < \epsilon$, such that

$$\sum_{i=0}^{n} p_i\left(\zeta_1, \ldots, \zeta_{r-1}, \zeta_{r+1}, \ldots, \zeta_k\right) \bar{z}_r^i = 0$$

In particular, we could choose all the ζ_i to have smaller magnitude than the respective $\hat{z}_i$. Therefore the point

$$\begin{bmatrix} \zeta_1 \\ \vdots \\ \zeta_{r-1} \\ \bar{z}_r \\ \zeta_{r+1} \\ \vdots \\ \zeta_k \end{bmatrix} \in \mathbf{C}^k$$

has $\|\cdot\|_\infty < \beta_p$, but is a root of $p(z)$. This contradicts the definition of β_p, hence this situation cannot occur.

Therefore, we have shown that if $\hat{z}$ is a minimizing solution, ie, $\|\hat{z}\|_\infty = \beta_p$ where $\beta_p = \min\{\|z\|_\infty : p(z) = 0\}$, then we may as well assume that each of the components of $\hat{z}$ has magnitude equal to β_p. ♯

This is the lemma necessary to prove that the lower bound is an equality.

Theorem 4.5 *Let $\mathbf{\Delta}$ be a given block structure, and let $\mathcal{Q}$ be defined as in section 3.1. Then for every matrix M of appropriate dimensions,*

$$\max_{Q \in \mathcal{Q}} \rho(QM) = \mu(M)$$

Proof: This is obvious if $\mu(M) = 0$, so we will assume that $\mu(M) > 0$. Let $\Delta \in \mathbf{\Delta}$ be a minimizing solution, so $\det(I + M\Delta) = 0$, and $\bar{\sigma}(\Delta) = \frac{1}{\mu(M)}$. Do a singular value decomposition on each block that makes up Δ. This gives $U, V \in \mathcal{Q}$, and a diagonal $\hat{\Sigma} \in \mathbf{\Delta}$, such that

$$\det\left(I + MU\hat{\Sigma}V^*\right) = 0$$

Since $\hat{\Sigma} \in \mathbf{\Delta}$ and is diagonal, it appears as

$$\hat{\Sigma} = \operatorname{diag}\left[\hat{\delta}_1 I_{r_1}, \ldots, \hat{\delta}_s I_{r_s}, \hat{\alpha}_1, \ldots, \hat{\alpha}_w\right]$$

for some complex numbers $\hat{\delta}_i$ and $\hat{\alpha}_j$, and $w = \sum_{j=1}^{f} m_j$. (recall the j'th full block is $m_j \times m_j$, hence each full block contributes m_j of the α's)

Consider $s + w$ complex **variables**, $z_1, \ldots, z_{s+w}$. Define a variable Σ by

$$\Sigma = \operatorname{diag}\left[z_1 I_{r_1}, \ldots, z_s I_{r_s}, z_{s+1}, \ldots, z_{s+w}\right]$$

Then $\det(I + MU\Sigma V^*)$ is a **polynomial** on $\mathbf{C}^{s+w}$, since the determinant involves only multiplications and additions. By hypothesis, a minimum norm root of this polynomial has an infinity norm (as defined above) of $\frac{1}{\mu(M)} =: \gamma$. Let $\bar{\Sigma}$ be the minimizing root with all components of equal magnitude, namely γ. Then we can write $\bar{\Sigma} = \gamma\Phi$ for some $\Phi \in \mathcal{Q}$. This gives

$$\det(I + \gamma MU\Phi V^*) = 0.$$

Obviously $\rho(MU\Phi V^*) \geq \mu(M)$, and the product $U\Phi V^* \in \mathcal{Q}$, so we are done. $\sharp$

5 Preliminaries for study of upper bound

The next major undertaking is a careful study of the upper bound: its computational properties, and the relation between μ and the upper bound. The purpose of this section is to collect some mathematical facts that we will need. All of the upcoming material appeared first in [Doy], although the theorems for the upper bound there are less general. Here we generalize the theorems in [Doy] to include block structures with repeated scalar blocks. Initially, we will focus on the $\bar{\sigma}(DMD^{-1})$ upper bound and begin by reparametrizing it.

5.1 Reparametrization of the upper bound

For the sake of computation, and proving some theorems, we must eliminate a degree of freedom present in the D's as they are defined now. From now on, we will assume that there is always at least 1 full uncertainty block, so that $f \geq 1$. The case with $s \geq 2$ and $f = 0$ is handled separately in section 12.1.

First, note that for any nonzero $\alpha \in \mathbf{C}$, and any $D \in \mathcal{D}$,

$$\bar{\sigma}\left(DMD^{-1}\right) = \bar{\sigma}\left((\alpha D)\, M\, (\alpha D)^{-1}\right). \tag{5.1}$$

Hence, in calculating the infimum, we can use this scaling, and without loss in generality, always assume that $d_f = 1$. Since we will have occasion to use it again though, we will now refer to the original set $\mathcal{D}$ as defined in (3.7) as $\mathcal{D}_g$.

In addition, we may assume that the other d_i are positive, and the D_i are positive definite. To see this, take $D \in \mathcal{D}$ and do a polar decomposition, $D = UP$ with U unitary and $P = P^* > 0$. Obviously

$$\bar{\sigma}\left(DMD^{-1}\right) = \bar{\sigma}\left(UPMP^{-1}U^*\right) = \bar{\sigma}\left(PMP^{-1}\right) \tag{5.2}$$

by the unitary invariance of $\bar{\sigma}$. Hence for any $D \in \mathcal{D}$, there is a positive definite, hermitian $D_H \in \mathcal{D}$ that achieves the same $\bar{\sigma}$. Therefore, the following definition for $\mathcal{D}_p$

$$\mathcal{D}_p = \left\{\text{diag}\ \left[D_1, \ldots, D_s, d_1 I_{m_1}, \ldots\, d_{f-1} I_{m_{f-1}}, I_{m_f}\right] : D_i = D_i^* \in \mathbf{C}^{r_i \times r_i} > 0, d_i > 0\right\} \tag{5.3}$$

leaves the infimum the same. **Note that implicitly, the last block has $d_f = 1$ as we indicated above.**

We do one further reparametrization via logarithms. Recall that

$$\left\{e^W : W = W^* \in \mathbf{C}^{m\times m}\right\} = \left\{D : D = D^* \in \mathbf{C}^{m\times m},\ \text{positive definite}\ \right\} \tag{5.4}$$

This simply says that the set of exponentials of all hermitian matrices is equal to the set of positive definite, hermitian matrices. The obvious block diagonal version of this fact allows us to redefine $\mathcal{D}$ as

$$\mathcal{D} := \left\{ \text{diag} \left[D_1, \ldots, D_s, d_1 I_{m_1}, \ldots, d_{f-1} I_{m_{f-1}}, 0_{m_f} \right] : D_i = D_i^* \in \mathbf{C}^{r_i \times r_i}, d_i \in \mathbf{R} \right\} \quad (5.5)$$

and the upper bound as

$$\mu_{\Delta}(M) \leq \inf_{D \in \mathcal{D}} \bar{\sigma}\left(e^{D} M e^{-D}\right). \quad (5.6)$$

We note that $\mathcal{D}$ is a finite dimensional, **real** (scalar multiplication must be real) vector space.

5.2 Convexity of the Upper Bound

In this section, we prove that the reparametrized upper bound is convex in the variable D. Therefore, any local minimum is also global minimum. Hence gradient optimization methods, which can yield local minima, can be used to nonconservatively compute the upper bound for μ. The first proof of this can be found in [SafD]. Here, we take an approach from [ChuD].

Definition 5.1 *Let* $\mathbf{X}$ *be a vector space. A function* $f : \mathbf{X} \to \mathbf{R}$ *is* **convex** *if for every* $x, y \in \mathbf{X}$, $\lambda \in [0,1]$

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

The next lemma gives a sufficient condition for a continuous function to be convex. It is fairly intuitive and is taken from [ChuD]. The proof is in the appendix.

Lemma 5.2 *Let* $f : \mathbf{R} \to \mathbf{R}$ *be a continuous function, and suppose for each* $t_o \in \mathbf{R}$, *there exists a twice differentiable function* $g_{t_o} : \mathbf{R} \to \mathbf{R}$, *such that* $f(t_o) = g_{t_o}(t_o)$, $f(t) \geq g_{t_o}(t)$ *for all* $t \in \mathbf{R}$ *and* $\left.\frac{d^2 g_{t_o}}{dt^2}\right|_{t=t_o} \geq 0$. *Then* f *is a convex function.*

We apply this to our situation.

Lemma 5.3 *For every* $D \in \mathcal{D}$, *the function* $f : \mathbf{R} \to \mathbf{R}$, $f(t) := \bar{\sigma}\left(e^{Dt} M e^{-Dt}\right)$ *is convex.*

Proof: We just need to verify the hypothesis of the lemma. Let t_o be given and let u_{t_o} and v_{t_o} be complex unit vectors, of appropriate dimensions such that

$$u_{t_o}^* e^{Dt_o} M e^{-Dt_o} v_{t_o} = \bar{\sigma}\left(e^{Dt_o} M e^{-Dt_o}\right).$$

For later use, we will let σ_o denote $\bar{\sigma}\left(e^{Dt_o} M e^{-Dt_o}\right)$, and $M_o := e^{Dt_o} M e^{-Dt_o}$.

Define $g_{t_o} : \mathbf{R} \to \mathbf{R}$ by

$$g_{t_o}(t) = \mathrm{Re}\left[u_{t_o}^* e^{Dt} M e^{-Dt} v_{t_o}\right] \tag{5.7}$$

Obviously, $f(t_o) = g_{t_o}(t_o)$, and for all $t \in \mathbf{R}$, $f(t) \geq g_{t_o}(t)$. Differentiating (5.7) twice gives

$$\left.\frac{d^2 g_{t_o}}{dt^2}\right|_{t=t_o} = \begin{bmatrix} u_{t_o}^* D & v_{t_o}^* D \end{bmatrix} \begin{bmatrix} \sigma_o I & -M_o^* \\ -M_o & \sigma_o I \end{bmatrix} \begin{bmatrix} Du_{t_o} \\ Dv_{t_o} \end{bmatrix} \tag{5.8}$$

Recall that $\bar{\sigma}(M_o) = \sigma_o$, hence the matrix in (5.8) is positive semidefinite, and therefore $\left.\frac{d^2 g_{t_o}}{dt^2}\right|_{t=t_o} \geq 0$. By Lemma 5.2, f is convex. $\sharp$

Trivially, we wrap this all up with

Lemma 5.4 *Consider the function $h : \mathcal{D} \to \mathbf{R}$, $h(D) = \bar{\sigma}\left(e^D M e^{-D}\right)$. Then h is convex.*

Proof: Let D_1 and D_2 be arbitrary elements in D, and let $\lambda \in [0,1]$. We need to show that

$$h((1-\lambda)D_1 + \lambda D_2) \leq (1-\lambda)h(D_1) + \lambda h(D_2)$$

Define $f : \mathbf{R} \to \mathbf{R}$ by $f(t) := h((1-t)D_1 + tD_2) = \bar{\sigma}\left[e^{\tilde{D}t}\left(e^{D_1} M e^{-D_1}\right) e^{-\tilde{D}t}\right]$ where $\tilde{D}$ is defined $\tilde{D} := D_2 - D_1$. Now, f is convex by Lemma 5.3, therefore for every $t \in [0,1]$

$$f(t) \leq (1-t)f(0) + tf(1) \tag{5.9}$$

Note that $f(0) = h(D_1)$ and $f(1) = h(D_2)$. Therefore, setting $t = \lambda$ in (5.9), we have

$$h((1-\lambda)D_1 + \lambda D_2) \leq (1-\lambda)h(D_1) + \lambda h(D_2) \tag{5.10}$$

as desired. $\sharp$

5.3 Directional derivatives of coalesced singular values

The minimization problem for the upper bound is discussed here. We calculate the first derivatives of singular values of $e^{Dt}Me^{-Dt}$ for given D in $\mathcal{D}$. The resulting formula will be used in section 7 to find a $D \in \mathcal{D}$ such that for $t > 0$, sufficiently small, $\bar{\sigma}\left(e^{D}Me^{-D}\right) < \bar{\sigma}(M)$, in other words, a descent direction for $\bar{\sigma}$. Iterating on this is a method to calculate the upper bound. In general, the minimization for the upper bound will drive the top singular values together, since we are minimizing a "max" function. Therefore, we must carry out the derivative calculations for coalesced singular values (ie. multiplicity greater than 1). Derivatives of distinct singular values are just special cases of the following results.

A result from perturbation theory, ([Kat] for the theory, [FreLC] and [Doy] for this application) that we will use freely is that if $T:\mathbf{R} \rightarrow \mathbf{C}^{n\times n}$ is an analytic function mapping the real line into hermitian matrices, then there exist analytic matrices $U(\cdot)$, and $\Lambda(\cdot)$, such that for all t, $U(t) \in \mathbf{C}^{n\times n}, U^*(t)U(t) = I$, $\Lambda(t) \in \mathbf{R}^{n\times n}$, $\Lambda(t)$ diagonal, and

$$T(t)U(t) = U(t)\Lambda(t). \tag{5.11}$$

In other words, the eigenvalues of an analytic hermitian matrix are analytic, and there is a choice of orthogonal analytic eigenvectors as well. We use this result to derive an expression for the derivatives of nonzero singular values of an analytic matrix.

Let $W:\mathbf{R} \rightarrow \mathbf{C}^{n\times m}$ be an analytic function of the real variable t. Suppose σ is a nonzero singular value of $W(0)$ with multiplicity r. Then σ^2 is a eigenvalue of $W(0)W^*(0)$, also with multiplicity r. Hence, there are analytic functions $U_a(\cdot), U_b(\cdot), \Sigma_a(\cdot)$, and $\Lambda_b(\cdot), \epsilon > 0$, such that for all $t \in (-\epsilon,\epsilon), U_a(t) \in \mathbf{C}^{n\times r}, U_b(t) \in \mathbf{C}^{n\times(n-r)}, \Sigma_a(t) \in \mathbf{R}^{r\times r}, \Lambda_b(t) \in \mathbf{R}^{(n-r)\times(n-r)}$ with both Σ_a and Λ_b diagonal and nonnegative for $t \in (-\epsilon,\epsilon)$. At $t = 0$, $\Sigma_a(0) = \sigma I_r$, and none of the diagonal entries of $\Lambda_b(0)$ are equal to σ^2. We also have that for all $t \in (-\epsilon,\epsilon)$

$$\begin{bmatrix} U_a^*(t) \\ U_b^*(t) \end{bmatrix} \begin{bmatrix} U_a(t) & U_b(t) \end{bmatrix} = I_{n\times n} \tag{5.12}$$

and

$$W(t)W^*(t) = U_a(t)\Sigma_a^2(t)U_a^*(t) + U_b(t)\Lambda_b(t)U_b^*(t) \tag{5.13}$$

We want to calculate the derivatives (at $t = 0$) of the r singular values which are coalesced at σ at $t = 0$. Of course, these are just the diagonal entries of $\dot{\Sigma}_a$, which itself is diagonal. Roughly speaking, we will differentiate (5.13) to get an explicit formula for $\dot{\Sigma}_a$.

Dropping the explicit t dependence, and post-multiplying (5.13) by $U_a(t)$ we have

$$WW^*U_a = U_a\Sigma_a^2 \tag{5.14}$$

Differentiating this gives

$$\dot{W}W^*U_a + W\dot{W}^*U_a + WW^*\dot{U}_a = \dot{U}_a\Sigma_a^2 + U_a\dot{\Sigma}_a\Sigma_a + U_a\Sigma_a\dot{\Sigma}_a$$

Premultiply this by U_a^*, and evaluate at $t = 0$. Recall that at $t = 0, \Sigma_a = \sigma I_r$. Hence, at $t = 0$

$$U_a^*\dot{W}W^*U_a + U_a^*W\dot{W}^*U_a + \sigma^2 U_a^*\dot{U}_a = \sigma^2 U_a^*\dot{U}_a + 2\sigma\dot{\Sigma}_a$$

Two terms cancel, and since $\sigma \neq 0$ by assumption, we are left with

$$\dot{\Sigma}_a = \frac{1}{2\sigma}U_a^*\left(\dot{W}W^* + W\dot{W}^*\right)U_a \tag{5.15}$$

Actual computation of the derivatives requires one additional computation. Consider a singular value decomposition of $W(0)$,

$$W(0) = \sigma U_1 V_1^* + U_2\Sigma_2 V_2^* \tag{5.16}$$

Since the singular vectors associated with repeated singular values are not unique, U_1 need not be equal to $U_a(0)$. But, both have orthogonal columns, and they span the same subspace in $\mathbf{C}^n$, therefore, there is a unitary matrix $K \in \mathbf{C}^{r\times r}$ such that

$$U_a(0) = U_1 K \tag{5.17}$$

Substituting (5.16) and (5.17) into (5.15) gives

$$K\dot{\Sigma}_a K^* = \frac{1}{2}\left(U_1^*\dot{W}V_1 + V_1^*\dot{W}^*U_1\right) \tag{5.18}$$

Since K is unitary, this is a similarity transformation, hence *the derivatives of the r singular values coalesced at σ are the eigenvalues of*

$$\frac{1}{2}\left(U_1^*\dot{W}V_1 + V_1^*\dot{W}^*U_1\right)$$

Let us do the above calculations for the special case we need.

Theorem 5.5 *Suppose $W(t)$ is of the form $e^{Dt}Me^{-Dt}$ where $D \in \mathcal{D}$ and M is given. Obviously $W(0) = M$ and $\dot{W}(0) = DM - MD$. Hence if*

$$W(0) = M = \sigma U_1 V_1^* + U_2\Sigma_2 V_2^* \tag{5.19}$$

then the derivatives of the clustered singular values at σ are the eigenvalues of

$$\sigma U_1^* D U_1 - \sigma V_1^* D V_1 \tag{5.20}$$

In particular, let $\lambda_1, \lambda_2, \ldots, \lambda_r$ be the eigenvalues of $U_1^*DU_1 - V_1^*DV_1$. They are real because this matrix is hermitian. At a nonzero value of t, the r singular values that were σ at $t = 0$ satisfy

$$\sigma_i(t) = \sigma(1 + \lambda_i t) + g_i(t) \tag{5.21}$$

where $\lim_{t\to 0}\frac{g_i(t)}{t} = 0$.

Hence if we can find a $D \in \mathcal{D}$ **with all the eigenvalues of** $U_1^* D U_1 - V_1^* D V_1$ **negative,** then by moving a small amount in that direction, all of the singular values in the cluster will be reduced.

After reviewing some results from convex analysis in the next section, we will address the problem of finding a $D \in \mathcal{D}$ such that for small t, all of the singular values in the cluster are reduced. As we have shown here, this is equivalent to finding a $D \in \mathcal{D}$ such that all the eigenvalues of $U_1^* D U_1 - V_1^* D V_1$ are negative.

5.4 Convexity

This section is devoted to some simple results from convex analysis which will subsequently be used to find $D \in \mathcal{D}$ such that all the eigenvalues of $U^* D U - V^* D V$ are positive. This gives a descent direction for $\bar{\sigma}\left(e^D M e^{-D}\right)$. All of the results here are from [Roc].

Let $\mathbf{X}$ be a real, finite dimensional vector space, with inner product $\langle \cdot, \cdot \rangle : \mathbf{X} \times \mathbf{X} \to \mathbf{R}$, and let $\mathcal{V}$ be a compact subset of $\mathbf{X}$. The main question this section addresses is "does there exist a point $\hat{x} \in \mathbf{X}$ and $\beta > 0$ such that $\min_{y \in \mathcal{V}} \langle \hat{x}, y \rangle \geq \beta$?"

The following definitions and results are standard.

Definition 5.6 *A subset $\mathcal{V} \subset \mathbf{X}$ is* **convex** *if $\lambda u + (1 - \lambda) v \in \mathcal{V}$ for every $u, v \in \mathcal{V}$ and $\lambda \in [0, 1]$.*

Definition 5.7 *For a subset $\mathcal{V} \subset \mathbf{X}$ the* **convex hull** *of $\mathcal{V}$, $\mathrm{co}(\mathcal{V})$ is the smallest convex set containing $\mathcal{V}$:*

$$\mathrm{co}(\mathcal{V}) = \bigcap_{\substack{\mathcal{F} \supset \mathcal{V} \\ \mathcal{F} \text{convex}}} \mathcal{F} \tag{5.22}$$

Lemma 5.8 *For all $\mathcal{V} \subset \mathbf{X}$, $\mathrm{co}(\mathcal{V})$ is convex. If $\mathcal{V}$ is convex, then $\mathrm{co}(\mathcal{V}) = \mathcal{V}$. If $\mathcal{V}$ is compact, then $\mathrm{co}(\mathcal{V})$ is compact.*

Lemma 5.9 *The convex hull of $\mathcal{V} \subset \mathbf{X}$ is all finite convex combinations of points in $\mathcal{V}$. That is*

$$\mathrm{co}(\mathcal{V}) = \left\{ \sum_{i=1}^{m} \alpha_i x_i : m \in \mathbf{N}, \alpha_i \in [0, 1], \sum_{i=1}^{m} \alpha_i = 1, \, x_i \in \mathcal{V} \right\} \tag{5.23}$$

Lemma 5.10 *Let $\mathcal{V}$ be a compact subset of $\mathbf{X}$. Then there is a unique point $\bar{x} \in \mathrm{co}(\mathcal{V})$ such that $\|\bar{x}\| = \min\{\|y\| : y \in \mathrm{co}(\mathcal{V})\}$. When clear, we denote this as $\bar{x} = \min(\mathrm{co}\mathcal{V})$.*

Lemma 5.11 *Let $\mathcal{V}$ be a compact subset of* $\mathbf{X}$, *and let $\bar{x} \in \operatorname{co}(\mathcal{V})$ be the unique minimizer desribed above, ie.* $\|\bar{x}\| = \min\{\|y\| : y \in \operatorname{co}(\mathcal{V})\}$. *For any $z \in \operatorname{co}(\mathcal{V})$*, $\langle z, \bar{x}\rangle \geq \|\bar{x}\|^2$.

Lemma 5.12 *Let $x \in \mathbf{X}$. If $\|x\| > \|\min(\operatorname{co}\mathcal{V})\|$, then there is a $y \in \mathcal{V}$ such that* $\langle x, y\rangle < \|x\|^2$.

These give rise to the main theorem.

Theorem 5.13 *Let $\mathcal{V}$ be a compact subset of* $\mathbf{X}$. *There exists $\hat{x} \in \mathbf{X}$ such that $\min_{y\in\mathcal{V}}\langle \hat{x}, y\rangle > 0$ if and only if $0 \notin \operatorname{co}(\mathcal{V})$.*

The minimum point of the convex hull of a set $\mathcal{V}$ can be found via an iterative algorithm, due to [Gil]. Important extensions of this are found in [Wol] and [Hau]. All the algorithms have one main computational requirement: for each $x \in \mathbf{X}$, we need to be able to generate a point $y_x \in \mathcal{V}$ such that

$$\langle x, y_x\rangle = \min_{y\in\mathcal{V}}\langle x, y\rangle \tag{5.24}$$

Note since $\mathcal{V}$ is closed, there always is such a y_x, though it may not be unique.

The algorithm from [Gil] is as follows: Define a sequence $\{x_i\}_{i=1}^{\infty}$ in the convex hull of $\mathcal{V}$ via the following rules:

a.1 Pick any point $x_1 \in \operatorname{co}\mathcal{V}$. In particular, x_1 can be any element of $\mathcal{V}$.

a.2 Given x_i, pick $y_i \in \mathcal{V}$ to minimize the inner product as above in equation (5.24).

a.3 Define $x_{i+1} = \min\operatorname{co}\{x_i, y_i\}$. Obviously, $x_{i+1} \in \operatorname{co}\mathcal{V}$. Return to **a.2.**

Hauser's algorithm [Hau] makes a more intelligent choice for x_{i+1}, using not only x_i and y_i, but past values of y_i as well. It is a generalization of Wolfe's algorithm [Wol] for polytopes. In any event,

Claim: The sequence $\{x_i\}$ converges to the minimum point in the convex hull of $\mathcal{V}$.

Proof of claim: Obviously, the sequence $\{x_i\}$ has $\|x_{i+1}\| \leq \|x_i\|$ for each i. Therefore both sequences $\{x_i\}$ and $\{y_i\}$ are bounded, hence we can choose a subsequence $\{n_k\}$ so that $x_{n_k} \xrightarrow{k} \bar{x}$ and $y_{n_k} \xrightarrow{k} \bar{y}$. Since both $\operatorname{co}\mathcal{V}$ and $\mathcal{V}$ are closed, we have $\bar{x} \in \operatorname{co}\mathcal{V}$ and $\bar{y} \in \mathcal{V}$. By continuity, and step [a.2] of the algorithm, it is easy to show that $\langle \bar{x}, \bar{y}\rangle = \min_{y\in\mathcal{V}}\langle \bar{x}, y\rangle$. Now suppose that $\bar{x} \neq \min(\operatorname{co}\mathcal{V})$. Since $\bar{x} \in \operatorname{co}\mathcal{V}$, we have by Lemmas 5.10 and 5.12 that

$$\langle \bar{x}, \bar{y}\rangle < \|\bar{x}\|^2. \tag{5.25}$$

Consequently, $\|\bar{x}\|$ is larger than $\|\min(\mathrm{co}\{\bar{x},\bar{y}\})\|$. Let $\epsilon > 0$ be the difference.

$$\epsilon := \|\bar{x}\| - \|\min(\mathrm{co}\{\bar{x},\bar{y}\})\| > 0 \tag{5.26}$$

Now the function $\min(\mathrm{co}\{\cdot,\cdot\}) : \mathbf{X} \times \mathbf{X} \to \mathbf{X}$ is a continuous function. Hence there is a integer K such for all $k > K$,

$$\|\min(\mathrm{co}\{x_{n_k}, y_{n_k}\})\| - \frac{\epsilon}{2} < \|\min(\mathrm{co}\{\bar{x},\bar{y}\})\| \tag{5.27}$$

This implies that for all $k > K$

$$\|\bar{x}\| > \|\min(\mathrm{co}\{x_{n_k}, y_{n_k}\})\| + \frac{\epsilon}{2} \tag{5.28}$$

which contradicts that the sequence $\{\|x_i\|\}$ is nonincreasing. Therefore $\bar{x} = \min(\mathrm{co}\mathcal{V})$.

Finally, it is an easy fact to show that if $\{z_k\}$ is a sequence in a compact, convex set, $\mathcal{G}$, with the norm satisfying $\|z_{k+1}\| \le \|z_k\|$ for all integers k, and there is a subsequence $\{z_{n_k}\}$ converging to $\min(\mathcal{G})$, then in fact, the sequence itself is convergent, with limit of course being $\min(\mathcal{G})$. Hence the sequence we generate, $\{x_i\}$, does indeed converge to the minimum point. $\sharp$

In the next section, we consider the problem of finding a matrix $D \in \mathcal{D}$, such that all of the eigenvalues of $U_1^* D U_1 - V_1^* D V_1$ are positive. Recall from Theorem 5.5, this is equivalent to finding a "descent direction" for the function $\bar{\sigma}\left(e^D M e^{-D}\right)$. This problem can be formulated naturally into a "minimum point in convex hull" formulation as we have covered here. We also show that finding a point $y_x \in \mathcal{V}$ that minimizes the inner product $\langle x, y\rangle$ can be cast as a hermitian eigenvalue problem.

6 Upper bound and the structured singular value

6.1 Finding Descent Directions for the Upper bound

6.1.1 Defining a Generalized Gradient Set

Our problem of finding a $D \in \mathcal{D}$ such that all the eigenvalues of $U^*DU - V^*DV$ are positive can be attacked using the convexity results from the previous section. The results are quite nice, and computationally tractable. The motivation comes from [Doy], though this section generalizes the results there.

We consider square matrices, $\mathbf{C}^{n\times n}$, and a compatible block structure $\mathbf{\Delta}$, with integers $r_1, \ldots, r_s$, $m_1, \ldots, m_f$ defining the dimensions of the blocks, as outlined in section 3.1. Define $\mathbf{X}$ to be the following set of block diagonal, hermitian matrices:

$$\mathbf{X} := \left\{ \text{ diag } [Z_1, \ldots, Z_s, z_1, \ldots, z_{f-1}] : Z_i = Z_i^* \in \mathbf{C}^{r_i \times r_i}, z_j \in \mathbf{R} \right\} \tag{6.1}$$

This is a real inner product space (of dimension $\sum_{i=1}^{s} r_i^2 + f - 1$) with inner product defined by

$$P, T \in \mathbf{X} \qquad \langle P, T \rangle := \operatorname{tr}(PT) \tag{6.2}$$

which, in terms of the blocks that make up P and L is just

$$\langle P, T \rangle = \sum_{i=1}^{s} \operatorname{tr}(P_i T_i) \; + \; \sum_{j=1}^{f-1} p_j t_j \tag{6.3}$$

Remark: When there are only full blocks, $s = 0$, then $\mathbf{X}$ is the set of $(f-1) \times (f-1)$, diagonal, real matrices, with the obvious inner product. In those instances, we will identify $\mathbf{X}$ with $\mathbf{R}^{f-1}$.

Recall the definition for $\mathcal{D}$ in (5.5). Let $D \in \mathcal{D}$ be given. Then D looks like

$$D = \text{ diag} \left[D_1, \ldots, D_s, \, d_1 I_{m_1}, \ldots, d_{f-1} I_{m_{f-1}}, 0_{m_f} \right] \tag{6.4}$$

where $D_i = D_i^* \in \mathbf{C}^{r_i \times r_i}$ and $d_j \in \mathbf{R}$. Associate to this $D \in \mathcal{D}$, a $\tilde{D} \in \mathbf{X}$ by setting

$$\tilde{D} \; = \; \text{diag } [D_1, \ldots, D_s, \, d_1, \ldots, d_{f-1}] \tag{6.5}$$

Note the natural one to one correspondence between the elements of $\mathcal{D}$ and $\mathbf{X}$.

Now, let $M \in \mathbf{C}^{n\times n}$ be given. If the maximum singular value of M, $\bar{\sigma}$, has multiplicity equal to r, then M is

$$M = \bar{\sigma} U V^* + U_2 \Sigma_2 V_2^* \tag{6.6}$$

where $U, V \in \mathbf{C}^{n\times r}$, $U^*U = V^*V = I_r$, $U_2, V_2 \in \mathbf{C}^{n\times(n-r)}$, $U_2^*U_2 = V_2^*V_2 = I_{(n-r)}$, and $\Sigma_2 \in \mathbf{R}^{(n-r)\times(n-r)}$ is diagonal, positive semidefinite, and none of its diagonal entries are equal to $\bar{\sigma}$.

Recall that we want to find a $D \in \mathcal{D}$ such that all the eigenvalues of $U^*DU - V^*DV$ are positive, or in other words, $\lambda_{\min} > 0$. Using Theorem 5.5, for such D, then with $t < 0$, sufficiently small in magnitude,

$$\bar{\sigma}\left(e^{Dt} M e^{-Dt}\right) < \bar{\sigma} \tag{6.7}$$

and hence computation of the $\inf\limits_{D\in\mathcal{D}} \bar{\sigma}\left(e^{D} M e^{-D}\right)$ depends on finding these D.

For notational purposes, partition U and V compatibly with $\mathbf{\Delta}$ as

$$U = \begin{bmatrix} A_1 \\ \vdots \\ A_s \\ E_1 \\ \vdots \\ E_f \end{bmatrix} \qquad V = \begin{bmatrix} B_1 \\ \vdots \\ B_s \\ F_1 \\ \vdots \\ F_f \end{bmatrix} \tag{6.8}$$

where $A_i, B_i \in \mathbf{C}^{r_i\times r}$, $E_i, F_i \in \mathbf{C}^{m_i\times r}$.

With this notation

$$U^*DU - V^*DV = \sum_{i=1}^{s} \left(A_i^* D_i A_i - B_i^* D_i B_i\right) + \sum_{j=1}^{f-1} d_j \left(E_j^* E_j - F_j^* F_j\right) \tag{6.9}$$

Therefore, since this matrix is hermitian, $\lambda_{\min}\left(U^*DU - V^*DV\right)$ is just

$$\lambda_{\min} = \min_{\substack{\eta\in\mathbf{C}^r \\ \|\eta\|=1}} \eta^* \left[\sum_{i=1}^{s} \left(A_i^* D_i A_i - B_i^* D_i B_i\right) + \sum_{j=1}^{f-1} d_j \left(E_j^* E_j - F_j^* F_j\right)\right] \eta \tag{6.10}$$

Exchanging the order of multiplication, and taking traces yields the equivalent form

$$\lambda_{\min} = \min_{\substack{\eta\in\mathbf{C}^r \\ \|\eta\|=1}} \left[\sum_{i=1}^{s} \operatorname{tr}\left[D_i \left(A_i \eta\eta^* A_i^* - B_i \eta\eta^* B_i^*\right)\right] + \sum_{j=1}^{f-1} d_j\, \eta^* \left(E_j^* E_j - F_j^* F_j\right) \eta\right]. \tag{6.11}$$

This can be rewritten using inner products as

$$\lambda_{\min} = \min_{\substack{\eta \in \mathbf{C}^r \\ \|\eta\|=1}} \langle \tilde{D}, P^\eta \rangle \tag{6.12}$$

where $P^\eta \in \mathbf{X}$ is defined by

$$\begin{aligned} P_i^\eta &:= A_i \eta \eta^* A_i^* - B_i \eta \eta^* B_i^* \\ p_j^\eta &:= \eta^* \left(E_j^* E_j - F_j^* F_j \right) \eta. \end{aligned} \tag{6.13}$$

Let $\nabla_M \subset \mathbf{X}$ be the set of all such P^η. That is

$$\nabla_M := \left\{ \text{diag } \left[P_1^\eta, \ldots, P_s^\eta, p_1^\eta, \ldots, p_{f-1}^\eta \right] : P_i^\eta, p_i^\eta \text{ as in (6.13)}, \eta \in \mathbf{C}^r, \|\eta\| = 1 \right\}. \tag{6.14}$$

Recall that when $r \geq 2$, the matrices U and V (which in turn define A, B, E and F above) are not unique. It is easy to verify that the set ∇_M **does not** depend on the particular choice.

Then, for a given $D \in \mathcal{D}$ (and corresponding $\tilde{D} \in \mathbf{X}$) we have

$$\lambda_{\min} (U^* D U - V^* D V) = \min_{P \in \nabla_M} \langle \tilde{D}, P \rangle. \tag{6.15}$$

Hence, it is the set ∇_M that determines whether or not there is a D that gives $\lambda_{\min} > 0$. The next theorem follows directly from equation (6.12) and Theorem 5.13.

Theorem 6.1 *There exists a* $D \in \mathcal{D}$ *such that* $\lambda_{\min} (U^* D U - V^* D V) > 0$ *if and only if* $0 \notin \text{co} (\nabla_M)$.

If $0 \in \text{co} (\nabla_M)$ then for every $D \in \mathcal{D}$, $\lambda_{\min} \leq 0$ and $\lambda_{\max} \geq 0$. Hence to first order, the maximum singular value either increases or stays the same (we are at a stationary point). By convexity of $\bar{\sigma} \left(e^D M e^{-D} \right)$, we see that we are at a global minimum. To summarize:

Theorem 6.2 $\bar{\sigma} (M) = \inf_{D \in \mathcal{D}} \bar{\sigma} \left(e^D M e^{-D} \right)$ *if and only if* $0 \in \text{co} (\nabla_M)$.

On occasion, we will abuse the notation ∇_M adopted above. When the matrix in question, in this case M, is clear from the context, we will drop the subscript and just write ∇.

Finally, we address the problem of computing the point of minimum norm in the convex hull of ∇_M. As mentioned in section 5.4, for each $\tilde{D} \in \mathbf{X}$, we need to be able to find a $P_{\tilde{D}} \in \nabla_M$ that achieves

$$\langle \tilde{D}, P_{\tilde{D}} \rangle = \min_{P \in \nabla_M} \langle \tilde{D}, P \rangle. \tag{6.16}$$

This is quite simple. Let $\tilde{D} \in \mathbf{X}$ be given, and let its components be D_i for $i = 1, \ldots, s$ and d_j for $j = 1, \ldots, f-1$. Then

$$\min_{P \in \nabla_M} \langle \tilde{D}, P \rangle = \min_{\substack{\eta \in \mathbf{C}^r \\ \|\eta\|=1}} \eta^* \underbrace{\left[\sum_{i=1}^{s} (A_i^* D_i A_i - B_i^* D_i B_i) + \sum_{j=1}^{f-1} d_j \left(E_j^* E_j - F_j^* F_j \right) \right]}_{W} \eta \qquad (6.17)$$

Obviously, the numerical value of this is just the minimum eigenvalue of the hermitian matrix W. Let $\eta_w \in \mathbf{C}^r$ be any unit length eigenvector associated with this eigenvalue, then

$$\arg \min_{P \in \nabla_M} \langle \tilde{D}, P \rangle = \operatorname{diag} \left[P_1^{\eta_w}, \ldots, P_s^{\eta_w}, p_1^{\eta_w}, \ldots, p_{f-1}^{\eta_w} \right] \in \nabla_M \qquad (6.18)$$

where the P's and p's are defined as

$$\begin{aligned} P_i^{\eta_w} &:= A_i \eta_w \eta_w^* A_i^* - B_i \eta_w \eta_w^* B_i^* \\ p_j^{\eta_w} &:= \eta_w^* \left(E_j^* E_j - F_j^* F_j \right) \eta_w. \end{aligned} \qquad (6.19)$$

for each i and j.

Using this formula, and the algorithm in [Hau], we can find the minimum point in the convex hull of ∇_M as desired.

6.1.2 A Property of ∇ when M is real

If the matrix M is real, then the minimum point in the convex hull of ∇ is real. We will prove this, and then see the implication it has on computing $\inf_{D \in \mathcal{D}} \bar{\sigma} \left(e^D M e^{-D} \right)$. Roughly speaking, each block of the optimal $D \in \mathcal{D}$ can be chosen to be real, symmetric.

Theorem 6.3 *If M is real, then for any block structure $\boldsymbol{\Delta}$, the minimum point in the convex hull of ∇_M is real.*

Proof: Since M is real, both U and V in the SVD of M may be taken as real. Now recall the algorithm to find $\min(\mathrm{co}\nabla_M)$ as described in the last chapter. We can pick x_1 to be any element of $\mathrm{co}\nabla_M$. If we choose an arbitrary **real** unit vector η_1, then our initial point x_1 is real. Obviously then, the point y_1 may be chosen real too. Simple induction gives that with this choice of x_1, the entire sequence $\{x_i\}$ is real. It converges to the minimum point, which therefore must be real. ♯

This leads to the next theorem.

Theorem 6.4 *Let $\mathcal{D}_R$ be the set of real, symmetric members of $\mathcal{D}$. If M is real, and the infimum*

$$\inf_{D_R \in \mathcal{D}_R} \bar{\sigma}\left(e^{D_R} M e^{-D_R}\right) \tag{6.20}$$

is achieved, then in fact

$$\inf_{D \in \mathcal{D}} \bar{\sigma}\left(e^{D} M e^{-D}\right) = \inf_{D_R \in \mathcal{D}_R} \bar{\sigma}\left(e^{D_R} M e^{-D_R}\right) \ \sharp \tag{6.21}$$

We make a conjecture that this is true even when the infimums are not achieved, but the details are not worked out here.

Conjecture 6.5 *Let $\mathcal{D}_R$ be the set of real, symmetric members of $\mathcal{D}$. If M is real, then*

$$\inf_{D \in \mathcal{D}} \bar{\sigma}\left(e^{D} M e^{-D}\right) = \inf_{D_R \in \mathcal{D}_R} \bar{\sigma}\left(e^{D_R} M e^{-D_R}\right) \tag{6.22}$$

6.2 When $\mu = \bar{\sigma}$

The results of this section relate the upper bound to μ.

As usual, let $\mathbf{\Delta}$ be a given structure, and let M be a given complex matrix. In the last section we showed that $\bar{\sigma}(M) = \inf \bar{\sigma}\left(e^{D} M e^{-D}\right)$ if and only if $0 \in \text{co}(\nabla_M)$. A natural question is: "When does $\bar{\sigma}(M) = \mu_{\Delta}(M)$?". The answer, which will link the upper bound and μ together, is the subject of the next theorem. Again, the set ∇ plays a crucial role.

Theorem 6.6 $\bar{\sigma}(M) = \mu_{\Delta}(M)$ *if and only if* $0 \in \nabla_M$.

Remark: This is exactly the result obtained in [Doy]. [Doy] however only considers structures with full blocks ($s = 0$). This section generalizes that result to structures with repeated scalar blocks as well.

Proof: For the proof, we follow the style of [Doy], and prove the equivalence of four statements:

1. $0 \in \nabla_M$
2. There exists $\eta \in \mathbf{C}^r$, $\|\eta\| = 1$ and $Q \in \mathcal{Q}$ such that $QU\eta = V\eta$
3. There exists $\xi \in \mathbf{C}^n$, $\|\xi\| = 1$ and $Q \in \mathcal{Q}$ such that $QM\xi = \bar{\sigma}\xi$
4. $\bar{\sigma}(M) = \mu_{\Delta}(M)$

$1 \to 2$: From the definition of ∇, (6.14), $0 \in \nabla_M$ implies that for some $\eta \in \mathbf{C}^r$, $\|\eta\| = 1$

$$\begin{array}{ll} A_i\eta\eta^* A_i{}^* - B_i\eta\eta^* B_i{}^* = 0 & i \le s \\ \eta^* \left(E_j{}^* E_j - F_j{}^* F_j\right) \eta = 0 & j \le f-1 \end{array} \tag{6.23}$$

Obviously, for $i \le s$, there is a phase $e^{j\theta_i}$ such that $e^{j\theta_i} A_i\eta = B_i\eta$. For $j \le f - 1$, $\|E_j\eta\| = \|F_j\eta\|$, so there exists a unitary matrix Q_j such that $Q_j E_j\eta = F_j\eta$. The only thing left is the last full block. Since $\|U\eta\| = \|V\eta\|$ we must have $\|E_f\eta\| = \|F_f\eta\|$. This gives a unitary matrix Q_f with $Q_f E_f\eta = F_f\eta$. Arranging the phases and Q's in a block diagonal fashion gives statement 2.

$2 \to 1$: This follows along the lines of $1 \to 2$.

$2 \to 3$: The matrix M has a SVD of $M = \bar{\sigma}UV^* + U_2\Sigma_2V_2^*$. Hence $QM(V\eta) = \bar{\sigma}QU\eta = \bar{\sigma}V\eta$. Defining $\xi = V\eta$ gives statement 3.

$3 \to 2$: A SVD of QM is

$$QM = \bar{\sigma}(QU)V^* + (QU_2)\Sigma_2V_2^* \tag{6.24}$$

If $QM\xi = \bar{\sigma}\xi$, then ξ must lie in the subspace spanned by the right singular vectors associated with $\bar{\sigma}$. Hence there is a vector η, satisfying $\xi = V\eta$. Obviously $\|\eta\| = 1$ and

$$QU\eta = QUV^*\xi = \frac{1}{\bar{\sigma}}QM\xi = \xi = V\eta. \tag{6.25}$$

$3 \to 4$: $QM\xi = \bar{\sigma}\xi$ implies that $\mu_\Delta(M) = \max_{Q\in\mathcal{Q}} \rho(QM) \ge \rho(QM) \ge \bar{\sigma}(M)$. However $\bar{\sigma}$ is always an upper bound for μ hence we must have equality.

$4 \to 3$: This is obvious by Theorem 4.5. $\left(\mu(M) = \max_{Q\in\mathcal{Q}} \rho(QM)\right)$ ♯

Theorem 6.6 is extremely important in determining when the upper bound gives μ. The idea is to find $D_o \in \mathcal{D}$ such that $0 \in \mathrm{co}(\nabla_{e^{D_o}Me^{-D_o}})$. This can in principle be done using a steepest descent method, and the facts about ∇ in section 6.1. Then, we know that

$$\mu(M) = \mu\left(e^{D_o}Me^{-D_o}\right) \le \inf_{D\in\mathcal{D}} \bar{\sigma}\left(e^{D}Me^{-D}\right) = \bar{\sigma}\left(e^{D_o}Me^{-D_o}\right). \tag{6.26}$$

If, in fact $0 \in \nabla_{e^{D_o}Me^{-D_o}}$, then by Theorem 6.6 we must have

$$\mu\left(e^{D_o}Me^{-D_o}\right) = \bar{\sigma}\left(e^{D_o}Me^{-D_o}\right) \tag{6.27}$$

so that

$$\mu(M) = \inf_{D\in\mathcal{D}} \bar{\sigma}\left(e^{D}Me^{-D}\right). \tag{6.28}$$

Therefore, if the block structure $\boldsymbol{\Delta}$ imparts the property on ∇ such that $0 \in \mathrm{co}(\nabla)$ implies $0 \in \nabla$, then we will always have $\mu_{\Delta}(M) = \inf_{D \in \mathcal{D}} \bar{\sigma}\left(e^{D} M e^{-D}\right)$.

A technical point we have not addressed is when the "inf" is not achieved. In that case the above reasoning cannot be used directly, since we never actually get $0 \in \mathrm{co}(\nabla)$. However, everything still works (this proof also rigorizes the above arguments):

Theorem 6.7 *If the block structure Δ has the property that $0 \in co(\nabla)$ always implies $0 \in \nabla$, then $\mu_{\Delta}(M) = \inf_{D \in \mathcal{D}} \bar{\sigma}\left(e^{D} M e^{-D}\right)$.*

Proof: Let $\beta = \inf \bar{\sigma}\left(e^{D} M e^{-D}\right)$. Let D_k be a sequence in $\mathcal{D}$ such that $\bar{\sigma}\left(e^{D_k} M e^{-D_k}\right)$ converges to β as $k \to \infty$. Denote $W_k = e^{D_k} M e^{-D_k}$. Since the sequence W_k is bounded, it has a convergent subsequence with limit W. Obviously, by continuity of $\bar{\sigma}$ and μ, $\bar{\sigma}(W) = \beta$ and $\mu(M) = \mu(W)$. We claim that $0 \in \mathrm{co}(\nabla_W)$. Suppose not, then there exist $\tilde{D} \in \mathcal{D}$ and $\epsilon > 0$ such that $\bar{\sigma}\left(e^{\tilde{D}} W e^{-\tilde{D}}\right) = \beta - \epsilon$. Choose k so that $\|W_k - W\| < \frac{\epsilon}{2\kappa(e^{\tilde{D}})}$, where $\kappa(\cdot)$ denotes condition number. Then

$$\|e^{\tilde{D}}(W_k - W)\, e^{-\tilde{D}}\| < \frac{\epsilon}{2} \tag{6.29}$$

which yields

$$\|e^{\tilde{D}} W_k e^{-\tilde{D}}\| < \beta - \frac{\epsilon}{2}. \tag{6.30}$$

This contradicts that β was the infimum, hence indeed $0 \in \mathrm{co}(\nabla_W)$. By hypothesis, this means $0 \in \nabla_W$ so by Theorem 6.6, $\mu(W) = \bar{\sigma}(W)$. Recalling continuity, we get $\mu_{\Delta}(M) = \beta$ as desired.$\sharp$

In the section to follow, we will determine some structures for which the hypothesis of Theorem 6.7 always holds. Therefore, for such structures, the upper bound will always equal μ.

To conclude this section, consider the minimization over the D's. Typically, since we are *minimizing* the *maximum* singular value, the top singular values tend to coalesce, so that at the minimum, the multiplicity of $\bar{\sigma}$ is greater than or equal to 2. This is typical of any "min max" problem. Suppose though, that at the minimum, $\bar{\sigma}(M)$ was distinct. Obviously, since we are at a minimum, we must have $0 \in \mathrm{co}(\nabla)$. But if the multiplicity of $\bar{\sigma}$ is only 1, then ∇ is **a single point**, and hence $\nabla = \{0\}$. This reasoning gives:

Corollary 6.8 *If, at the minimum of $\bar{\sigma}\left(e^{D} M e^{-D}\right)$, the maximum singular value has multiplicity of 1, then $\mu(M) = \min_{D \in \mathcal{D}} \bar{\sigma}\left(e^{D} M e^{-D}\right)$.*

7 Properties of the set ∇_M

With the machinery presented in the last section, we can now explore the relationship between the upper bound an μ for a variety of block structures.

7.1 Block structures with no *repeated scalar* blocks

We begin with block structures having no repeated scalar blocks, that is, when $s = 0$. All material here is taken from [Doy] and [MorD] and is included for completeness.

7.1.1 2 full blocks

The situation with two full blocks is relatively simple. Referring back to (6.13) and (6.14), we see that ∇ will always have the form

$$\nabla = \{\eta^* (E^*E - F^*F)\, \eta : \eta \in \mathbf{C}^r, \|\eta\| = 1\} \tag{7.1}$$

for some given $r > 0$ and $E, F \in \mathbf{C}^{m_1 \times r}$. Since $E^*E - F^*F$ is hermitian, ∇ is just a closed interval in the real line. Obviously, this is always convex, so if $0 \in \mathrm{co}\,(\nabla)$, we in fact have $0 \in \nabla$. Hence by theorem 6.7 we have:

Theorem 7.1 *If Δ consists of two full blocks $(s = 0, f = 2)$, then*

$$\mu_\Delta (M) = \inf_{D \in \mathcal{D}} \bar{\sigma} \left(e^D M e^{-D}\right). \tag{7.2}$$

Remark: The two block case was first solved in 1959 by Redheffer [Red]. His approach is quite different. Interestingly, it uses a form of Schauder's fixed point theorem, [DunS] and hence does not boil down to just simple linear algebra. Similarly, the method of proof here uses the analyticity of eigenvalues of an analytic matrix, which is also a nontrivial fact. It would be quite nice if simpler proofs existed, but none are known.

Also, this is a fairly simple thing to compute. Recall that for two full blocks, there is only **one** free parameter in the set $\mathcal{D}$, consequently, the computation is a one dimensional search on a convex function. The only drawback is that the cost evaluation is a $\bar{\sigma}$ evaluation, which while not exceedingly difficult, is nonetheless time consuming. Note that a search need not involve gradient calculations, hence the code can be quite simple.

7.1.2 4 full blocks

Consider the case when Δ consists of four 1×1 blocks, so $s = 0$, $f = 4$, and $m_j = 1$ for each j. Let a, b, and c be positive real numbers, d and f be complex numbers, and ψ_1 and ψ_2 be real angles. Define matrices $U, V \in \mathbf{C}^{4\times 2}$ by

$$U = \begin{bmatrix} a & 0 \\ b & b \\ c & jc \\ d & f \end{bmatrix} \quad , \quad V = \begin{bmatrix} 0 & a \\ b & -b \\ c & -jc \\ e^{j\psi_1} f & e^{j\psi_2} d \end{bmatrix} \tag{7.3}$$

For the time being, suppose that these are both unitary matrices, so that $U^*U = V^*V = I_2$. Later we will actually assign the correct values, but at the moment we just assume this is already done. Then define $M \in \mathbf{C}^{4\times 4}$ by

$$M := UV^* \tag{7.4}$$

With the assumptions of unitariness on U and V, (7.4) *is* a singular value decompostion of M. M has two singular values at 1, and two singular values at 0. With respect to the block structure Δ that we have defined, what properties does the set ∇_M have? In particular:

- is $0 \in \text{co}(\nabla_M)$? If so, then $\inf_{D\in\mathcal{D}} \bar{\sigma}\left(e^D M e^{-D}\right) = 1$, otherwise, it is less than 1.
- is $0 \in \nabla_M$? If so, then $\mu(M) = \bar{\sigma}(M) = 1$, otherwise it is less than 1.

Since the multiplicity of the maximum singular value is 2, we can parametrize all unit vectors in $\mathbf{C}^2$, and get a parametric representation of ∇_M. It is easy to see that any vector $\eta \in \mathbf{C}^2$, with $\|\eta\| = 1$ is of the form

$$\eta = \begin{bmatrix} e^{j\phi_1}\cos\theta \\ e^{j\phi_2}\sin\theta \end{bmatrix}$$

for some real ϕ_1, ϕ_2, and θ. As it turns out, ∇_M depends only on the difference $\phi_1 - \phi_2$, which we will denote as ϕ.

Simply plugging in for the definition of ∇_M from section 6.1.1, we get

$$\nabla_M = \left\{ \begin{bmatrix} a^2\left(\cos^2\theta - \sin^2\theta\right) \\ 4b^2 \sin\theta\cos\theta\cos\phi \\ 4c^2\sin\theta\cos\theta\sin\phi \end{bmatrix} \in \mathbf{R}^3 : \phi, \theta \in \mathbf{R} \right\} \subset \mathbf{R}^3 \tag{7.5}$$

It is apparent that $0 \notin \nabla_M$. That would require (from the first coordinate in (7.5)) that $\theta = \frac{2n+1}{4}\pi$, for some integer n. The second and third coordinates being zero would then require both $\cos\phi = 0$ and $\sin\phi = 0$, which is impossible. Hence $0 \notin \nabla_M$, and $\mu(M) < 1$.

On the other hand, setting $\theta = 0$, and then $\theta = \frac{\pi}{2}$, gives that both $[a^2\ 0\ 0]^T$ and $[-a^2\ 0\ 0]^T$ are elements of ∇_M. Consequently, $0 \in \mathrm{co}(\nabla_M)$. Therefore

$$\inf_{D \in \mathcal{D}} \bar{\sigma}\left(e^D M e^{-D}\right) = \bar{\sigma}(M) = 1$$

In order to complete the counterexample, we must choose the free variables so that U and V in (7.4) are unitary, as we said we could. In fact here, we will choose them so that ∇_M is the boundary of a ball in $\mathbf{R}^3$, centered at the origin. The radius happens to be $\frac{2}{3+\sqrt{3}}$. This particular choice of parameters was obtained via alot of algebra.

Set $\gamma = 3 + \sqrt{3}$ and $\beta = \sqrt{3} - 1$ and define

$$a = \sqrt{\frac{2}{\gamma}} \quad , \quad b = \frac{1}{\sqrt{\gamma}} \quad , \quad c = \frac{1}{\sqrt{\gamma}}$$

$$d = -\sqrt{\frac{\beta}{\gamma}} \quad , \quad f = (1+j)\sqrt{\frac{1}{\gamma\beta}}$$

$$\psi_1 = -\frac{\pi}{2} \quad , \quad \psi_2 = \pi$$

Some algebra later, we conclude that ∇_M is the set of all $x \in \mathbf{R}^3$, such that $\|x\| = \frac{2}{3+\sqrt{3}}$. Obviously, $0 \notin \nabla_M$, but $0 \in \mathrm{co}\nabla_M$. Extensive searching over the set $\mathcal{Q}$ in the lower bound formula (recall that while the lower bound is always μ, unfortunately, it is not a concave function, so gradient methods yield only local maxima) has revealed that for M defined above, $\mu(M)$ is approximately 0.874.

Therefore, for the 4 full block problem, as opposed to the 2 full block problem, in general, $\mu(M) \neq \inf_{D \in \mathcal{D}} \bar{\sigma}\left(e^D M e^{-D}\right)$. Since the **full** blocks in this counterexample are 1×1, they may be viewed as **repeated scalar** blocks as well. Therefore this counterexample proves that for every block structure $\boldsymbol{\Delta}$ satisfying $s + f \geq 4$, in general, we will have

$$\mu(M) \neq \inf_{D \in \mathcal{D}} \bar{\sigma}\left(e^D M e^{-D}\right).$$

7.1.3 3 Full Blocks

In view of the 2 previous sections, the only case with $s = 0$ that we don't know about is 3 full blocks. In this section, we will prove that indeed, ∇ is always convex, and hence for every matrix M, the infimum upper bound is equal to μ. Recall that if $\boldsymbol{\Delta}$ consists of 3 full blocks ($s=0, f=3$), then ∇ is of the form

$$\nabla = \left\{ \begin{bmatrix} \eta^* H_1 \eta \\ \eta^* H_2 \eta \end{bmatrix} \in \mathbf{R}^2 : \eta \in \mathbf{C}^r, \|\eta\| = 1 \right\} \subset \mathbf{R}^2 \tag{7.6}$$

for some integer r, and hermitian matrices H_1 and $H_2 \in \mathbf{C}^{r\times r}$. Obviously, if $r = 1$, then the set ∇ is a single point, so it is convex. The next 3 lemmas will show that for any positive r, this is also convex.

We begin with some notation from [Doy]. For any positive integer r, we define the sets $P^r := \{x \in \mathbf{C}^r : \|x\| = 1\}$ and $S^r := \{v \in \mathbf{R}^{r+1} : \|v\| = 1\}$. If $H_1, H_2, \ldots, H_q$ are hermitian matrices in $\mathbf{C}^{r\times r}$, we define a function $f_H : P^r \rightarrow \mathbf{R}^q$ by

$$f_H(\eta) := \begin{bmatrix} \eta^* H_1 \eta \\ \eta^* H_2 \eta \\ \vdots \\ \eta^* H_q \eta \end{bmatrix} \in \mathbf{R}^q \tag{7.7}$$

for each $\eta \in P^r$.

Lemma 7.2 *Let q be a positive integer. Let $a_i, c_i \in \mathbf{R}$, and $b_i \in \mathbf{C}$ for $i = 1, \ldots, q$. For each i, define a hermitian 2×2 matrix H_i by*

$$H_i := \begin{bmatrix} a_i & b_i \\ \bar{b}_i & c_i \end{bmatrix}$$

Then there exists a vector $d \in \mathbf{R}^q$ and a matrix $V \in \mathbf{R}^{q\times 3}$ such that

$$f_H\left(P^2\right) = \left\{d + Vu : u \in S^2\right\}.$$

where f_H is defined in (7.7).

Remark: In other words, the image of P^2 by f_H is the image of an affine linear map on the unit disk in $\mathbf{R}^3$.

Proof: First, we parametrize the unit ball in $\mathbf{C}^2$ as

$$\eta = \begin{bmatrix} e^{j\omega} \cos\theta \\ e^{j\psi} \sin\theta \end{bmatrix}$$

for some real ω, ψ, and θ. As it turns out, only on the difference $\omega - \psi$ is important, and we denote this as ϕ.

Then, for any one of the particular H_i, and for any $\eta \in \mathbf{C}^2$, with $\|\eta\| = 1$ we have

$$\begin{aligned}
\eta^* H \eta &= \begin{bmatrix} e^{-j\omega}\cos\theta & e^{-j\psi}\sin\theta \end{bmatrix} \begin{bmatrix} a & b \\ \bar{b} & c \end{bmatrix} \begin{bmatrix} e^{j\omega}\cos\theta \\ e^{j\psi}\sin\theta \end{bmatrix} \\
&= a\cos^2\theta + c\sin^2\theta + 2\left[\mathrm{Re}(b)\cos\phi + \mathrm{Im}(b)\sin\phi\right]\cos\theta\sin\theta \\
&= \tfrac{a+c}{2} + \tfrac{a-c}{2}\cos 2\theta + 2\left[\mathrm{Re}(b)\cos\phi + \mathrm{Im}(b)\sin\phi\right]\cos\theta\sin\theta \qquad (7.8) \\
&= \tfrac{a+c}{2} + \begin{bmatrix} \tfrac{a-c}{2} & \mathrm{Re}(b) & \mathrm{Im}(b) \end{bmatrix} \begin{bmatrix} \cos 2\theta \\ 2\cos\phi\cos\theta\sin\theta \\ 2\sin\phi\cos\theta\sin\theta \end{bmatrix}
\end{aligned}$$

Note that the vector

$$\begin{bmatrix} \cos 2\theta \\ 2\cos\phi\cos\theta\sin\theta \\ 2\sin\phi\cos\theta\sin\theta \end{bmatrix}$$

is a parametrization of S^2. Hence setting $d_i := \frac{a_i+c_i}{2}$ and the i'th row of V, v_i, to

$$v_i := \begin{bmatrix} \frac{a_i-c_i}{2} & \mathrm{Re}(b_i) & \mathrm{Im}(b_i) \end{bmatrix}$$

proves the lemma. ♯

Lemma 7.3 *Let* $\bar{d} \in \mathbf{R}^2$ *and* $V \in \mathbf{R}^{2\times 3}$. *Then the set* $G_{\bar{d},V} := \left\{\bar{d} + Vu : u \in S^2\right\}$ *is convex.*

Proof: Let $u_1, u_2 \in S^2$ and let $\lambda \in [0,1]$. Obviously

$$\lambda\left(\bar{d} + Vu_1\right) + (1-\lambda)\left(\bar{d} + Vu_2\right) = \bar{d} + V\left(\lambda u_1 + (1-\lambda)u_2\right).$$

Now $\|\lambda u_1 + (1-\lambda)u_2\| \le 1$. If it is equal to 1, we are done. Otherwise, we can add to it a vector w in the null space of V (note because of the dimensions, V always has a nontrivial nullspace) so that $u_3 := \lambda u_1 + (1-\lambda)u_2 + w \in S^2$. Then

$$\lambda\left(\bar{d} + Vu_1\right) + (1-\lambda)\left(\bar{d} + Vu_2\right) = \bar{d} + Vu_3 \in G_{\bar{d},V}. ♯$$

Hence, for $q = 2$ and $r = 2$, the set $f\left(P^2\right) \in \mathbf{R}^2$ is convex. For a block structure with $s = 0, f = 3$, the set ∇ is always of the form $f\left(P^r\right) \in \mathbf{R}^2$ (ie. $q = 2$). Recall though, that in our application, r is the multiplicity of the maximum singular value. Concievably, this can be anything, hence we need to generalize the above reasoning for $r > 2$. This is easy.

Lemma 7.4 *Let r be any positive integer. Let $H_1, H_2 \in \mathbf{C}^{r\times r}$ be hermitian matrices. Then the set*

$$f_H(P^r) = \left\{ \begin{bmatrix} \eta^* H_1 \eta \\ \eta^* H_2 \eta \end{bmatrix} \in \mathbf{R}^2 : \eta \in \mathbf{C}^r, \|\eta\| = 1 \right\} \tag{7.9}$$

is convex.

Proof: Let η_1, η_2 be unit vectors in $\mathbf{C}^r$ and let $\lambda \in [0,1]$. With f_H defined in (7.9), we need to find a $\eta_3 \in P^r$ such that

$$f_H(\eta_3) = \lambda f_H(\eta_1) + (1-\lambda) f_H(\eta_2).$$

Without loss of generality, suppose $\eta_1 \neq \eta_2$. Choose orthogonal vectors $x, y \in P^r$ that span the same two-dimensional subspace as that spanned by η_1 and η_2. Define two hermitian matrices $\hat{H}_1$ and $\hat{H}_2 \in \mathbf{C}^{2\times 2}$ by

$$\hat{H}_i := \begin{bmatrix} x^* \\ y^* \end{bmatrix} H_i \; [x\; y].$$

Using these two matrices, and the definition of f in (7.7), we can naturally define a function $f_{\hat{H}}: P^2 \to \mathbf{R}^2$. From Lemma 7.3, we know that the set $f_{\hat{H}}(P^2)$ is convex. Since x and y are orthogonal, the matrix $[x\,y] \in \mathbf{C}^{r\times 2}$ is unitary, and there are vectors $\zeta_1, \zeta_2 \in P^2$ such that $\eta_i = [x\,y]\,\zeta_i$ for each $i = 1, 2$. Therefore, for each i, $f_H(\eta_i) = f_{\hat{H}}(\zeta_i)$. Now by convexity of $f_{\hat{H}}(P^2)$, there is a $\zeta_3 \in P^2$ such that

$$\lambda f_{\hat{H}}(\zeta_1) + (1-\lambda) f_{\hat{H}}(\zeta_2) = f_{\hat{H}}(\zeta_3)$$

Let $\eta_3 \in P^r$ be defined by $\eta_3 := [x\,y]\,\zeta_3$. Note that $f_H(\eta_3) = f_{\hat{H}}(\zeta_3)$. Therefore, $\lambda f_H(\eta_1) + (1-\lambda) f_H(\eta_2) = f_H(\eta_3)$, so that $f_H(P^r)$ is indeed convex as claimed.$\sharp$

7.1.4 Summary for block structures with $s = 0$

The last three sections have shown the well known results for block structures with only full blocks. These results were alluded to in the top row of the table from section 3.1. As we noted in section 7.1.2, the counterexample for 4 full blocks is also a counterexample for other block structures, since the full blocks in the example were 1×1 and could be viewed a repeated scalar blocks as well.

It is not know what the worst ratio of μ over the upper bound can be. The 4 block counterexample in this section has a ratio of approximately .874. Extensive computational experience has failed to reveal another example which is worse, even for much higher number of blocks. There has not yet been a physically motivated example where the ratio was more than .98.

The situation when there are also repeated scalar blocks, $s \neq 0$, has not been been studied as extensively. One of these structures is the topic of the next section.

7.2 Block structures with $s \neq 0$

As we saw in the last section, when $s = 0$ and $f \leq 3$ (3 or less full blocks), the set ∇_M is itself convex. Therefore for that block structure, $\mu(M) = \inf_{D \in \mathcal{D}} \bar{\sigma}\left(e^D M e^{-D}\right)$. In addition, there also exist 4 block examples where $0 \in \mathrm{co}(\nabla)$ but $0 \notin \nabla$. Of course, by the previous results, $\mu \neq \inf_{D \in \mathcal{D}} \bar{\sigma}$ in those instances.

Until now, the case of *repeated scalar* ($s \neq 0$) blocks has not been investigated. In section 6.1.1, we defined the correct ∇_M set to obtain descent directions for $\bar{\sigma}\left(e^D M e^{-D}\right)$ when repeated scalar blocks are part of the block structure. Then in section 6.2, we showed that $0 \in \nabla_M$ if and only if $\mu(M) = \bar{\sigma}(M)$, a result previously known for the case of all full blocks. In this section, we continue with structures having repeated scalar blocks, in particular, we consider a block structure of *one* repeated scalar block, and *one* full block. Recall the definition of ∇_M, equation (6.14). With this structure, the set ∇_M will always be of the form

$$\nabla = \{A\eta\eta^* A^* - B\eta\eta^* B^* : \eta \in \mathbf{C}^r, \|\eta\| = 1\} \tag{7.10}$$

for some given $r > 0$ and $A, B \in \mathbf{C}^{r_1 \times r}$. It is easy to see that in general, ∇ is not convex. For instance, take $A = I$ and $B = 0$. Then ∇ is all norm 1 dyads, but in general, a convex combination of norm 1 dyads **is not** a norm 1 dyad, so ∇ is not convex. However the following (which is all we need) is always true.

Theorem 7.5 *Let ∇ be defined as in (7.10). If $0 \in \mathrm{co}(\nabla)$, then $0 \in \nabla$.*

Proof: Suppose that $0 \in \mathrm{co}(\nabla)$. Then, for some integer p, there exist nonnegative α_i, $i = 1, 2, \ldots, p$ with $\sum_{i=1}^{p} \alpha_i = 1$ and vectors η_i, $i = 1, 2, \ldots, p$ with $\|\eta_i\| = 1$ such that

$$\sum_{i=1}^{p} \alpha_i \left(A\eta_i\eta_i^* A^* - B\eta_i\eta_i^* B^*\right) = 0 \tag{7.11}$$

which is rewritten as

$$A\left(\sum_{i=1}^{p} \alpha_i \eta_i \eta_i^*\right) A^* = B\left(\sum_{i=1}^{p} \alpha_i \eta_i \eta_i^*\right) B^* \tag{7.12}$$

Since the α_i are nonnegative, and not all 0, the dyad summation in (7.12) is a positive semidefinite matrix that is not zero. Let $X^{\frac{1}{2}}$ be its hermitian, positive semidefinite square root. Therefore

$$AX^{\frac{1}{2}}X^{\frac{1}{2}}A^* = BX^{\frac{1}{2}}X^{\frac{1}{2}}B^* \tag{7.13}$$

Hence, there is a unitary matrix V such that

$$AX^{\frac{1}{2}} = BX^{\frac{1}{2}}V \tag{7.14}$$

Let v be an eigenvector of V (with eigenvalue $e^{j\theta}$) such that $X^{\frac{1}{2}}v \neq 0$, and define $u := X^{\frac{1}{2}}v$. Note that u is nonzero. This gives

$$Au = e^{j\theta}Bu \tag{7.15}$$

which implies that $0 \in \nabla$. $\sharp$

What implication does this have? Obviously for this structure, $\mu(M) = \inf_{D \in \mathcal{D}} \bar{\sigma}\left(DMD^{-1}\right)$. Precisely, let M be a given matrix, partitioned as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

and suppose the dimensions are $M_{11} \in \mathbf{C}^{n \times n}$ and $M_{22} \in \mathbf{C}^{m_2 \times m_1}$. Define $\mathbf{\Delta}$ as

$$\mathbf{\Delta} = \left\{ \text{diag } [\delta_1 I_n , \Delta_2] : \delta_1 \in \mathbf{C} , \Delta_2 \in \mathbf{C}^{m_1 \times m_2} \right\}$$

Then

$$\mu_{\mathbf{\Delta}}(M) = \inf_{\substack{D \in \mathbf{C}^{n \times n} \\ D \text{ invertible}}} \bar{\sigma} \begin{bmatrix} DM_{11}D^{-1} & DM_{12} \\ M_{21}D^{-1} & M_{22} \end{bmatrix}$$

8 Linear Fractional Transformations

8.1 Introduction

Using only the definition of μ, we can prove some rather simple theorems about a class of general linear feedback loops called **Linear Fractional Transformations**. To introduce these, consider a complex matrix M partitioned as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \tag{8.1}$$

and suppose there is a defined block structure $\boldsymbol{\Delta}$ which is compatible in size with M_{11}. For $\Delta \in \boldsymbol{\Delta}$, consider the following loop equations,

$$\begin{aligned} e &= M_{21}w + M_{22}d \\ z &= M_{11}w + M_{12}d \\ w &= \Delta z \end{aligned} \tag{8.2}$$

This set of equations (8.2) is called **well posed** if for any vector d, there exist unique vectors w, z, and e satisfying the loop equations. It is easy to see that the set of equations is well posed if and only if the inverse of $I - M_{11}\Delta$ exists. If not, then depending on d and M, there is either no solution to the loop equations, or there are an infinite number of solutions. When the inverse does indeed exist, we have $e = F_u(M,\Delta)d$ where

$$F_u(M,\Delta) = M_{22} + M_{21}\Delta(I - M_{11}\Delta)^{-1}M_{12} \tag{8.3}$$

$F_u(M,\Delta)$ is called a **Linear Fractional Transformation** on M by Δ, and in a feedback diagram appears as:

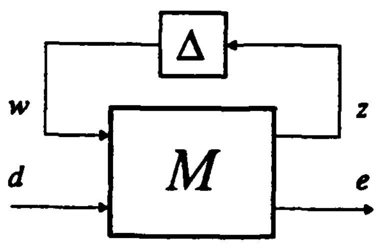


Figure 8.1 Linear Fractional Transformation

From a system point of view, we interpret vector d as the "disturbance", and e is the "error", whereas vectors z and w are internal variables. M_{22} is the nominal map between the disturbance and error, and Δ represents unknown quantities, called perturbations, which affect the map in a known way-namely through M_{12}, M_{21}, M_{11}, and the formula F_u.

The subscript u on F_u pertains to the "upper" loop of M is closed by Δ. An analogous formula describes $F_l(M,\Delta)$, which is the resulting matrix obtained by closing the "lower" loop of M (assuming the dimensions are ok and the implied inverse exists).

The constant matrix problem that we would like to solve is:

- determine whether the LFT is well posed for all Δ in some prescribed subset $\mathbf{\Omega} \subset \mathbf{\Delta}$ and,
- if so, then determine how "large" $F_u(M, \Delta)$ can get for Δ in $\mathbf{\Omega}$.

The next section has three simple theorems which answer this problem.

8.2 Well Posedness and Performance for Constant LFT's

One appealing use of μ is to determine the well posedness of a linear fractional transformation on a structured Δ, and to determine how "big" the linear fractional transformation can get. As we will see, μ answers these questions. Of course, using the results here will require that we can compute μ.

Consider a complex matrix M partitioned as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \tag{8.4}$$

and suppose there are two defined block structures $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$ which are compatible in size with M_{11} and M_{22} respectively. Define a third structure $\tilde{\mathbf{\Delta}}$ as

$$\tilde{\mathbf{\Delta}} = \left\{ \begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix} : \Delta_1 \in \mathbf{\Delta}_1, \Delta_2 \in \mathbf{\Delta}_2 \right\} \tag{8.5}$$

Now we have three structures with which we may compute μ with respect to. The notation we will use to keep track of this is as follows: $\mu_1(\cdot)$ is with respect to $\mathbf{\Delta}_1$, $\mu_2(\cdot)$ is with respect to $\mathbf{\Delta}_2$, : $\mu_{1,2}(\cdot)$ is with respect to $\tilde{\mathbf{\Delta}}$. In view of this, $\mu_1(M_{11})$, $\mu_2(M_{22})$ and $\mu_{1,2}(M)$ all make sense, though for instance, $\mu_1(M)$ does not.

The first theorem addresses the well posedness of the LFT $F_u(M, \Delta_1)$, and is nothing more than a restatement of the definition of μ.

Theorem 8.1 *Let $\beta > 0$. The LFT is well posed for all $\Delta_1 \in \frac{1}{\beta}\mathbf{B\Delta}$ if and only if $\mu_1(M_{11}) < \beta$.*

Note that the $\leq$ and $<$ signs can be exchanged and the theorem is still true. An imprecise but important notion to get from this is that *the minimum amount of structured feedback necessary to cause a loop to be ill posed is inversely proportional to μ of the open loop.*

As the "perturbation" Δ_1 deviates from zero, the matrix relating d to e deviates from M_{22}. Using the quantity $\mu_{1,2}(M)$, we can bookkeep what happens to $\mu_2(F_u(M,\Delta_1))$ as follows:

Theorem 8.2 (Robust Performance:constant) *Let $\beta > 0$. Then $\mu_{1,2}(M) < \beta$ if and only if $\mu_1(M_{11}) < \beta$, and for all $\Delta_1 \in \mathbf{B\Delta}_1$, $\mu_2(F_u(M,\Delta_1)) < \beta$.*

Proof:

$\leftarrow$ Let $\Delta_i \in \mathbf{\Delta}_i$ be given, with $\bar{\sigma}(\Delta_i) \leq \frac{1}{\beta}$, and define $\Delta = \text{diag}\,[\Delta_1, \Delta_2]$. Obviously $\Delta \in \tilde{\mathbf{\Delta}}$. Now

$$\det(I - M\Delta) = \det \begin{bmatrix} I - M_{11}\Delta_1 & -M_{12}\Delta_2 \\ -M_{21}\Delta_1 & I - M_{22}\Delta_2 \end{bmatrix} \tag{8.6}$$

By hypothesis $I - M_{11}\Delta_1$ is invertible, hence $\det(I - M\Delta)$ becomes

$$\det(I - M_{11}\Delta_1)\det\left(I - M_{22}\Delta_2 - M_{21}\Delta_1(I - M_{11}\Delta_1)^{-1} M_{12}\Delta_2\right)$$

Collecting the Δ_2 terms leaves

$$\det(I - M\Delta) = \det(I - M_{11}\Delta_1)\det(I - F_u(M,\Delta_1)\Delta_2)$$

We also have $\mu_2(F_u(M,\Delta_1)) < \beta$, so, since $\bar{\sigma}(\Delta_2) \leq \frac{1}{\beta}$, the quantity $I - F_u(M,\Delta_1)\Delta_2$ must be nonsingular. Therefore $I - M\Delta$ is nonsingular, so $\mu_{1,2}(M) < \beta$.

$\rightarrow$ Basically, you just reverse the argument above, but we include this for completeness. Again let $\Delta_1 \in \mathbf{\Delta}_1$ and $\Delta_2 \in \mathbf{\Delta}_2$ be given, with $\bar{\sigma}(\Delta_i) \leq \frac{1}{\beta}$, and define $\Delta = \text{diag}\,[\Delta_1, \Delta_2]$. By hypothesis, we know that $I - M\Delta$ is nonsingular. It is easy to verify from the definition of μ that (always)

$$\mu_{1,2}(M) \geq \max\{\mu_1(M_{11})\ ,\ \mu_2(M_{22})\}$$

so we also have $\mu_1(M_{11}) < \beta$, which gives that $I - M_{11}\Delta_1$ is nonsingular too. Therefore

$$\det(I - M_{11}\Delta_1)\det(I - F_u(M,\Delta_1)\Delta_2) = \det(I - M\Delta) \neq 0$$

Obviously, $I - F_u(M,\Delta_1)\Delta_2$ is nonsingular.♯

An identical proof switches the $\leq$ and $<$ signs:

Theorem 8.3 *Let $\beta > 0$. Then $\mu_{1,2}(M) \leq \beta$ if and only if $\mu_1(M_{11}) \leq \beta$, and for all $\Delta_1 \in \mathbf{\Delta}_1$, with $\bar{\sigma}(\Delta_1) < \frac{1}{\beta}$ $\mu_2(F_u(M,\Delta_1)) \leq \beta$.*

Roughly speaking, we have a test that determines if for all $\bar{\sigma}(\Delta_1) \le \frac{1}{\beta}$, the quantity $\mu_2(F_u(M,\Delta_1))$ stays bounded by β. Since both $\rho(\cdot)$ and $\bar{\sigma}(\cdot)$ are special cases of μ, by the appropriate choice of the set $\mathbf{\Delta}_2$, either $\rho(F_u(M,\Delta_1))$ or $\bar{\sigma}(F_u(M,\Delta_1))$ could be "watched". Of course for different choices of $\mathbf{\Delta}_2$, the theorem gives information about $\mu_2(F_u(M,\Delta_1))$.

Note that in this test, *the bound we get on the performance is dependent on the bound we set on the perturbation*, namely they are reciprocals. For other values, we must scale M and recompute. Specifically, for $\alpha \ge 0$, define $\mathbf{M}_\alpha$ as

$$\mathbf{M}_\alpha = \begin{bmatrix} M_{11} & M_{12} \\ \alpha M_{21} & \alpha M_{22} \end{bmatrix} \tag{8.7}$$

Some simple facts about $\mathbf{M}_\alpha$:

- If $\alpha = 0$ then $\mu_{1,2}(\mathbf{M}_\alpha) = \mu_1(M_{11})$
- For any $\Delta_1 \in \mathbf{\Delta}_1$, $F_u(\mathbf{M}_\alpha, \Delta_1) = \alpha F_u(M,\Delta_1)$ (as long as the inverse exists)
- $\max\{\mu_1(M_{11}) \ , \ \alpha\mu_2(M_{22})\} \le \mu_{1,2}(\mathbf{M}_\alpha) \le \max\{1,\alpha\}\,\mu_{1,2}(M)$
- $\mu_{1,2}(\mathbf{M}_\alpha)$ is a continuous, nondecreasing function of α

Let $\gamma > \mu_1(M_{11})$ be given, and define

$$\alpha_\gamma = \max_{\alpha>0}\{\alpha : \mu_{1,2}(\mathbf{M}_\alpha) = \gamma\} \tag{8.8}$$

These lead to the following variant of Theorems 8.2 and 8.3;

Theorem 8.4 (Worst Case:constant) *Let $\gamma > \mu_1(M_{11})$ be given, and α_γ be computed from (8.8). Then*

$$\sup_{\Delta_1 \in \frac{1}{\gamma}\mathbf{B\Delta}_1} \mu_2(F_u(M,\Delta_1)) = \frac{\gamma}{\alpha_\gamma} \tag{8.9}$$

Remark: The basic idea of the theorem is this: find the largest α such that for all $\Delta_1 \in \frac{1}{\beta}\mathbf{B\Delta}$, we still get $\mu_2(F_u(M,\Delta_1)) < \frac{\beta}{\alpha}$. By the 2nd fact above, this is the same as: find the largest α such that for all $\bar{\sigma}(\Delta_1) \le \frac{1}{\beta}$, $\mu_2(\mathbf{M}_\alpha) < \beta$. This test we can do, by applying Theorem 8.2 on $\mathbf{M}_\alpha$, which then gives the result.

Proof: Since $\gamma > \mu_1(M_{11})$, the left hand side of (8.9) is always well defined. By definition of α_γ, we know that $\mu_{1,2}\left(\mathbf{M}_{\alpha_\gamma}\right) = \gamma$, and for any $\epsilon > 0$, $\mu_{1,2}\left(\mathbf{M}_{\alpha_\gamma+\epsilon}\right) > \gamma$. Applying Theorem 8.3 with $\beta = \gamma$ gives

$$\text{if}\quad \Delta_1 \in \mathbf{\Delta}_1, \bar{\sigma}(\Delta_1) < \frac{1}{\gamma} \quad, \text{ then } \quad \mu_2(F_u(M,\Delta_1)) \leq \frac{\gamma}{\alpha_\gamma}$$

Since $F_u(M,\Delta_1)$ is well defined and continuous on $\left\{\Delta_1 : \bar{\sigma}(\Delta_1) \leq \frac{1}{\gamma}\right\}$, we have

$$\sup_{\Delta_1 \in \frac{1}{\gamma}\mathbf{B\Delta}_1} \mu_2(F_u(M,\Delta_1)) \leq \frac{\gamma}{\alpha_\gamma}$$

Suppose it is truly less. Then for some $\epsilon > 0$

$$\sup_{\Delta_1 \in \frac{1}{\gamma}\mathbf{B\Delta}_1} \mu_2(F_u(M,\Delta_1)) = \frac{\gamma}{\alpha_\gamma + \epsilon}$$

which implies that $\mu_{1,2}\left(\mathbf{M}_{\alpha_\gamma+\epsilon}\right) \leq \gamma$, a contradiction of the definition of α_γ. $\sharp$

Corollary 8.5 *If $\mu_{1,2}(\mathbf{M}_\alpha)$ is a* **increasing** *function of α (not just nondecreasing) then*

$$\sup_{\bar{\sigma}(\Delta_1) \leq \frac{1}{\mu(\mathbf{M}_\alpha)}} \mu_2(F_u(M,\Delta_1)) = \frac{\mu_{1,2}(\mathbf{M}_\alpha)}{\alpha}.$$

Finally, we state a maximum modulus like result for μ. The proof uses Theorem 4.5 from the previous section, along with ideas similar to the ones here.

Theorem 8.6 (Maximum modulus: LFT) *Let M be given as in (8.4), along with two block structures $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$. Suppose that $\mu_1(M_{11}) < 1$. Then*

$$\max_{\Delta_1 \in \mathbf{B\Delta}_1} \mu_2(F_u(M,\Delta_1)) = \max_{Q_1 \in \mathcal{Q}_1} \mu_2(F_u(M,Q_1)) \tag{8.10}$$

Remark: In light of this, any μ test with at least one repeated scalar block can always be reduced to a one dimensional search of μ tests without that block. A similar result to Theorem 8.6 is in [BoyD]. They show for that any H bounded and analytic on $|z| \leq 1$, the function $k(z) := \mu(H(z))$ is subharmonic.

Finally, we note that Theorems 8.1 through 8.4, along with corollary 8.5 and Theorem 8.6 have obvious analogs dealing with the behavior of $F_l(M,\Delta)$, under structured perturbations. In this section, all of the results were stated and proven for $F_u(M,\Delta)$. Throughout this thesis, we will use the result of either type without special mention.

8.3 Examples of LFT's

8.3.1 Transfer functions as LFT's

Consider a stable, discrete time, linear system

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k \end{aligned} \tag{8.11}$$

with transfer function $G(z) = D + C(zI - A)^{-1}B$ (n states, and for simplicity, we assume that this has m inputs and outputs, though everything that follows holds for nonsquare plants also). The infinity norm of G is defined as

$$\|G\|_\infty = \sup_{\substack{z \in \mathbf{C} \\ |z| \geq 1}} \bar{\sigma}(G(z))$$

which is equivalent to

$$\|G\|_\infty = \sup_{\substack{\delta \in \mathbf{C} \\ |\delta| \leq 1}} \bar{\sigma}\left(D + \delta C(I - \delta A)^{-1}B\right) \tag{8.12}$$

Define $\mathbf{\Delta}_1 = \{\delta I_n : \delta \in \mathbf{C}\}$, $\mathbf{\Delta}_2 = \mathbf{C}^{m \times m}$ and

$$M := \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathbf{R}^{(n+m)\times(n+m)}. \tag{8.13}$$

In μ notation, we can write (8.12) as

$$\|G\|_\infty = \sup_{\Delta_1 \in \mathbf{B}\mathbf{\Delta}_1} \mu_2\left(F_u(M, \Delta_1)\right) \tag{8.14}$$

because the block structure $\mathbf{\Delta}_2$ implies that $\mu_2(\cdot) = \bar{\sigma}(\cdot)$, and $\mathbf{\Delta}_1$ has been defined to represent the $\mathcal{Z}$-transform variable. Applying theorem 8.2, with $\beta = 1$, gives

$$\|G\|_\infty < 1 \quad \text{iff} \quad \mu_{1,2}(M) < 1. \tag{8.15}$$

In view of the result in section 7.2, actually $\|G\|_\infty < 1$ if and only if there exists a coordinate transformation $T \in \mathbf{C}^{n \times n}$ such that

$$\bar{\sigma}\left(\begin{bmatrix} TAT^{-1} & TB \\ CT^{-1} & D \end{bmatrix}\right) < 1.$$

Hence, we have an algorithm for *generating all stable rational transfer functions that have* $\|\cdot\|_\infty < 1$. Simply choose any matrix M so that $\bar{\sigma}(M) < 1$ and partition M as shown above. Then G will be stable, and have norm less than one, and all stable rational $G(z)$, with $\|G\|_\infty < 1$ can be generated in this fashion.

This result can also be shown using results from dissipative systems, and linear quadratic optimal control theory (with nondefinite cost functions). In fact, if $\|G\|_\infty \leq 1$, then solving one Riccati equation yields a $T \in \mathbf{C}^{n\times n}$ such that

$$\bar{\sigma}\left(\begin{bmatrix} TAT^{-1} & TB \\ CT^{-1} & D \end{bmatrix}\right) = 1.$$

The details of this calculation are interesting, and follow straightforwardly from the results in [Wil]. We do not include them here because the Riccati solution has the undesirable property that n of the singular values will be coalesced at $\bar{\sigma} = 1$. This seems to limit the usefulness of the Riccati solution as a viable computational alternative to gradient searching along the "full" D directions.

In this example, the "perturbation" is the *repeated scalar* block, and for the $\|\cdot\|_\infty$ norm, it **must** correspond to the unit disk. Using theorem 8.2 with β equal to 1, we can only check if $\|G\|_\infty$ is less than 1. For other values, we must scale G and recompute, using Theorem 8.4. Namely, define $\bar{\alpha}$ as

$$\bar{\alpha} = \max_{\alpha>0}\left\{\alpha : \inf_T \bar{\sigma}\left(\begin{bmatrix} TAT^{-1} & TB \\ \alpha CT^{-1} & \alpha D \end{bmatrix}\right) = 1\right\} \tag{8.16}$$

Then the worst case theorem, Theorem 8.4 (with $\gamma = 1$) gives

$$\|G\|_\infty = \frac{1}{\bar{\alpha}} \tag{8.17}$$

8.3.2 Keeping LFT's large

Just as μ can be used to determine how big the maximum singular value (or spectral radius) of an LFT can get, we can also use it to determine if the minimum singular value will remain bounded away from 0 (and, of course, the minimum eigenvalue too). Of course, the motivation of the LFT as a "perturbed disturbance to error" is no longer applicable, but this problem is interesting in its own right. The key to all this is that the inverse of an LFT, $F_l(M,\Delta)$, is itself an LFT, on the same Δ, but with a different known matrix, M_I.

This section will present these types of results. All are obtained from the well known "matrix inversion lemma", which we review for completeness. We begin with a lemma that is fundamental to the matrix inversion lemma.

Lemma 8.7 *Let A, B, C, and D be complex matrices, $A \in \mathbf{C}^{n\times n}, B \in \mathbf{C}^{n\times m}, C \in \mathbf{C}^{m\times m}, D \in \mathbf{C}^{m\times n}$. Suppose that A and C are each invertible. Then $A + BCD$ is invertible if and only if $C^{-1} + DA^{-1}B$ is invertible.*

Proof: Taking determinants, we get

$$\begin{aligned}\det(A+BCD) &= \det A \det(I+A^{-1}BCD)\\ &= \det A \det(I+DA^{-1}BC)\\ &= \det A \det(C^{-1}+DA^{-1}B) \det C. \sharp\end{aligned}$$

In order to evaluate how small things actually get, we need the matrix inversion lemma.

Lemma 8.8 *Suppose A, B, C, and D are given as in lemma 5.6. If A, C, and $A+BCD$ are invertible, then*

$$(A+BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1}+DA^{-1}B\right)^{-1}DA^{-1}$$

Proof: By lemma 8.7, $C^{-1}+DA^{-1}B$ is invertible – the result follows by verification. $\sharp$

Now, let M be given, partitioned in a 2×2 fashion as in (8.4), and let $\mathbf{\Delta}_2$ be a given structure, compatible with M_{22}. Suppose M_{11} is square, hence $F_l(M,\Delta)$ is square too. Under what conditions is $F_l(M,\Delta_2)$ invertible for all $\Delta_2 \in \mathbf{\Delta}_2$, with $\bar{\sigma}(\Delta) < \frac{1}{\beta}$?

First, we require that it be well defined for all such Δ_2, so we need $\mu_2(M_{22}) \leq \beta$. This guarantees that $I + M_{22}\Delta_2$ will be invertible. Second, it is obvious that M_{11} needs to be invertible, otherwise the LFT is not invertible even for $\Delta_2 = 0$.

Theorem 8.9 *Let M be given, with the following assumptions: M_{11} is square and invertible, and $\mu_2(M_{22}) < \beta$. Then for all $\Delta_2 \in \frac{1}{\beta}\mathbf{B\Delta}_2$ the LFT $F_l(M,\Delta_2)$ is invertible if and only if $\mu_2\left(M_{22} - M_{21}M_{11}^{-1}M_{12}\right) < \beta$.*

Proof: Since M_{11} is invertible, and $\mu_2(M_{22}) < \beta$, and $\bar{\sigma}(\Delta_2) \leq \frac{1}{\beta}$, we can apply Lemma 8.7 to determine the invertibility of

$$\underbrace{M_{11}}_{A} + \underbrace{M_{12}\Delta_2}_{B}\underbrace{(I-M_{22}\Delta_2)^{-1}}_{C}\underbrace{M_{21}}_{D}.$$

This is invertible if and only if

$$I - M_{22}\Delta_2 + M_{21}M_{11}^{-1}M_{12}\Delta_2 \tag{8.18}$$

is invertible. Recall the definition of μ. The quantity in (8.18) is invertible for all $\Delta_2 \in \mathbf{\Delta}_2$ with $\bar{\sigma}(\Delta_2) \leq \frac{1}{\beta}$ if and only if $\mu_2\left(M_{22} - M_{21}M_{11}^{-1}M_{12}\right) < \beta. \sharp$

Now apply the matrix inversion lemma to get an expression for $[F_l(M,\Delta_2)]^{-1}$. If $F_l(M,\Delta_2)$ is invertible, then

$$[F_l(M,\Delta_2)]^{-1} = M_{11}^{-1} - M_{11}^{-1}M_{12}\Delta_2\left[I - \left(M_{22} - M_{21}M_{11}^{-1}M_{12}\right)\Delta_2\right]^{-1} M_{21}M_{11}^{-1}$$

If we define a matrix M_I as

$$M_I := \begin{bmatrix} M_{11}^{-1} & -M_{11}^{-1}M_{12} \\ M_{21}M_{11}^{-1} & M_{22}-M_{21}M_{11}^{-1}M_{12} \end{bmatrix}$$

then

$$[F_l(M,\Delta_2)]^{-1} = F_l(M_I,\Delta_2)$$

Theorem 8.10 *Suppose that M_{11} is invertible, $\mu_2\left(M_{22} - M_{21}M_{11}^{-1}M_{12}\right) < \beta$, and $\mu_2(M_{22}) < \beta$. Then, in view of the discussion, the following equivalences make sense and are true:*

$$\min_{\Delta_2 \in \frac{1}{\beta}\mathbf{B\Delta_2}} \underline{\sigma}[F_l(M,\Delta_2)] > \frac{1}{\beta} \quad \leftrightarrow \quad \max_{\Delta_2 \in \frac{1}{\beta}\mathbf{B\Delta_2}} \bar{\sigma}(F_l(M_I,\Delta_2)) < \beta$$
$$\leftrightarrow \quad \mu_{\hat{\mathbf{\Delta}}}(M_I) < \beta$$

where $\hat{\mathbf{\Delta}} := \{\mathrm{diag}[\Delta,\Delta_2] : \Delta \in \mathbf{C}^{n\times n}, \Delta_2 \in \mathbf{\Delta}_2\}$. *(If we had wanted to keep track of the smallest magnitude eigenvalue, as opposed to the smallest singular value, then the top block of $\hat{\mathbf{\Delta}}$ would instead be a repeated scalar block)* ♯

8.4 Upper bound LFT results

Each of the Theorems 8.2 and 8.3 give necessary and sufficient conditions for some performance/robustness characteristic in terms of a μ evaluation. Looking back at these theorems, we see that the μ test always looks like "Is $\mu(M) < \beta$?" (or $\leq$). Hence, upper and lower bounds can be used in the following manner:

- an upper bound gives a sufficient condition for the robustness/performance characteristic of the theorem
- a lower bound gives a sufficient condition when the robustness/performance will not be met

Consequently, both are important. The upper bound will yield positive comments like "We are okay for perturbations up to this size, and maybe alot better", while the lower bound yields negative statements such as "There is a perturbation this size that will cause instability (or sufficient degradation in performance), and it might be worse".

The above comments apply for **any** upper and lower bound. In this section, we will concentrate on the additional information that is obtained in using the $\bar{\sigma}(DMD^{-1})$ upper bound. In other words, because of its structure, $\bar{\sigma}(DMD^{-1}) \leq \beta$ in general implies a great deal more than $\mu(M) \leq \beta$. One word before proceeding: we drop the exponential notation for D's, and revert back to (3.7) for the notation. Recall that the exponential parametrization was introduced in section 5 to allow simpler derivative formulas, which are implicit in the definition of ∇_M.

We begin with an obvious result: **Trivial upper bound lemma:** *Let $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$ be two given structures, and define a third, $\tilde{\mathbf{\Delta}} = \{\mathrm{diag}[\Delta_1, \Delta_2] : \Delta_i \in \mathbf{\Delta}_i\}$. Let M be a given matrix such that $\mu_{1,2}(M)$ makes sense, and suppose there is a function $\mu_{ub}(\cdot)$ that is a upper bound for $\mu_{1,2}(\cdot)$. If $\mu_{ub}(M) < \beta$, then*

$$\max_{\Delta_1 \in \frac{1}{\beta}\mathbf{B\Delta}_1} \mu_2(F_u(M, \Delta_1)) < \beta$$

Proof: This follows directly from the **constant matrix robust performance theorem**, Theorem 8.2.♯

The following theorem shows what additional information we get if the upper bound, $\mu_{ub}(\cdot)$, is in fact the $\bar{\sigma}(DMD^{-1})$ upper bound. As before, let $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$ be two given structures, and let $\tilde{\mathbf{\Delta}} = \{\mathrm{diag}[\Delta_1, \Delta_2] : \Delta_i \in \mathbf{\Delta}_i\}$. Similarly, let $\mathcal{D}_i$ be the appropriate D scaling sets for the two structures, and denote $\tilde{\mathcal{D}}$ as the obvious diagonal augmentation of these two sets.

Lemma 8.11 (Constant D lemma) *Let M be given as in the* **robust performance** *theorem, 8.2. Suppose there is a $\tilde{D} \in \tilde{\mathcal{D}}$ such that*

$$\bar{\sigma}\left(\tilde{D}M\tilde{D}^{-1}\right) < \beta$$

Then there exists a $D_2 \in \mathcal{D}_2$ such that

$$\max_{\Delta_1 \in \frac{1}{\beta}\mathbf{B\Delta}_1} \bar{\sigma}\left(D_2 F_u(M, \Delta_1) D_2^{-1}\right) < \beta$$

Remarks: Initially, one might guess that if we replace μ by the $\bar{\sigma}(DMD^{-1})$ upper bound in the robust performance theorem hypothesis, the resulting claim would just have μ replaced by $\bar{\sigma}(DMD^{-1})$. This lemma shows that we get quite abit more: If the $\bar{\sigma}(DMD^{-1})$

upper bound is less than β, this does not just imply that for all $\Delta_1 \in \mathbf{\Delta}_1$, with $\bar{\sigma}(\Delta_1) \leq \frac{1}{\beta}$, the upper bound of $F_u(M, \Delta_1)$ is less than β. It implies, instead, that this is indeed so, **but using only a single $D_2 \in \mathcal{D}_2$.**

Proof: The easiest method of proof is just to track the norms of the various vectors in the loop equations for the LFT. Let D_1 and D_2 be the separate parts of the $\tilde{D} \in \tilde{\mathcal{D}}$ that achieves $\bar{\sigma}\left(\tilde{D}M\tilde{D}^{-1}\right) < \beta$. Obviously, $\mu_1(M_{11}) < \beta$, so for any $\Delta_1 \in \mathbf{\Delta}_1$ with $\bar{\sigma}(\Delta_1) \leq \frac{1}{\beta}$, the two LFT's below are well posed, and from d to e are the same.

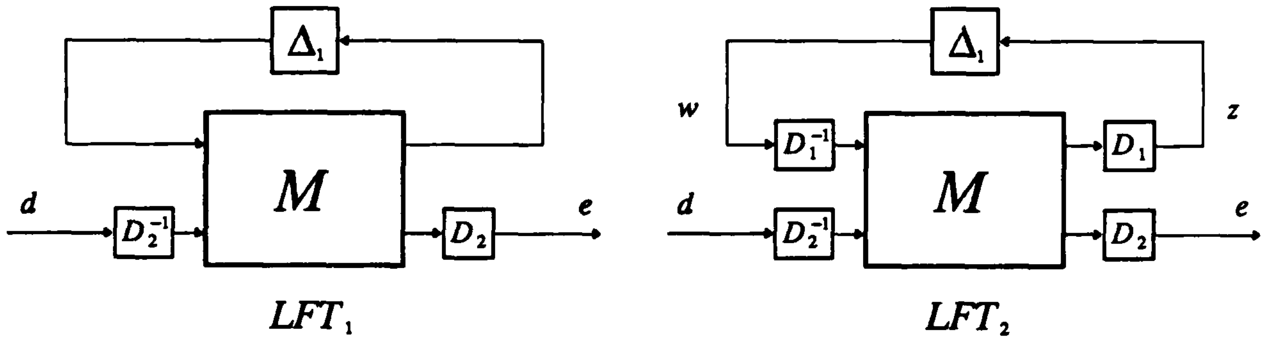


Figure 8.2 Diagram for Proof of Lemma 8.11

Let $d \neq 0$ be any given complex vector of appropriate dimension, and let e, w, and z be the unique solutions to the loop equations for LFT_2. By hypothesis, we have

$$\|z\|^2 + \|e\|^2 < \beta^2 \left(\|w\|^2 + \|d\|^2\right) \tag{8.19}$$

and since $\bar{\sigma}(\Delta_1) \leq \frac{1}{\beta}$

$$\|w\|^2 \leq \frac{1}{\beta^2} \|z\|^2 \tag{8.20}$$

Combining these gives that

$$\|e\|^2 < \beta^2 \|d\|^2. \tag{8.21}$$

Equation (8.21) also holds for LFT_1, since the map from d to e is the same for both LFTs. This implies that $\bar{\sigma}\left(D_2 F_u(M, \Delta_1) D_2^{-1}\right) < \beta$ as desired. ♯

An interesting question is "what is the optimal constant scaling that one can apply?" In particular, suppose $\mu_1(M_{11}) < 1$. Therefore, for all $\Delta_1 \in \mathbf{\Delta}_1$, with $\bar{\sigma}(\Delta_1) \leq 1$, the linear fractional transformation $F_u(M, \Delta_1)$ is defined. Can we compute the value of

$$\inf_{D_2 \in \mathcal{D}_2} \max_{\Delta_1 \in \mathbf{B\Delta}_1} \bar{\sigma}\left(D_2 F_u(M, \Delta_1) D_2^{-1}\right) \tag{8.22}$$

and also find a D_2 that achieves it? Towards answering this question, we have a simple lemma:

Lemma 8.12 *Let M, $\mathbf{\Delta}_1$, $\mathbf{\Delta}_2$, $\mathcal{D}_1$, and $\mathcal{D}_2$ be given as usual. Suppose that the $\mathbf{\Delta}_2$ structure has dimension $n_2 \times n_2$. Define an augmented structure $\hat{\mathbf{\Delta}}$ as*

$$\hat{\mathbf{\Delta}} := \left\{ \operatorname{diag}[\Delta_1, \Delta] : \Delta_1 \in \mathbf{\Delta}_1, \Delta \in \mathbf{C}^{n_2 \times n_2} \right\} \tag{8.23}$$

Note that $\hat{\mathbf{\Delta}}$ **is not** *$\mathbf{\Delta}_1$ augmented with $\mathbf{\Delta}_2$. It is $\mathbf{\Delta}_1$ augmented with an unstructured block the same size as $\mathbf{\Delta}_2$. Suppose that $\mu_1(M_{11}) < 1$. Then for $\alpha > 0$, and $D_2 \in \mathcal{D}_2$,*

$$\mu_{\hat{\Delta}} \begin{pmatrix} M_{11} & M_{12}D_2^{-1} \\ \alpha D_2 M_{21} & \alpha D_2 M_{22} D_2^{-1} \end{pmatrix} < 1 \tag{8.24}$$

if and only if

$$\max_{\Delta_1 \in \mathbf{B\Delta}_1} \bar{\sigma}\left(D_2 F_u(M, \Delta_1) D_2^{-1} \right) < \frac{1}{\alpha}. \tag{8.25}$$

Proof: Again, this follows directly from the definition of the structure $\hat{\mathbf{\Delta}}$, and the **robust performance theorem**, Theorem 8.2.♯

This allows easy proof of the **General optimal constant scaling** theorem:

Theorem 8.13 *Let M, $\mathbf{\Delta}_1$, $\mathbf{\Delta}_2$, $\mathcal{D}_1$, $\mathcal{D}_2$, and $\hat{\mathbf{\Delta}}$ be given as in Lemma 8.12. Suppose that $\mu_1(M_{11}) < 1$. Define γ by*

$$\gamma = \sup_{\alpha > 0} \left\{ \alpha : \inf_{D_2 \in \mathcal{D}_2} \mu_{\hat{\Delta}} \begin{pmatrix} M_{11} & M_{12}D_2^{-1} \\ \alpha D_2 M_{21} & \alpha D_2 M_{22} D_2^{-1} \end{pmatrix} < 1 \right\} \tag{8.26}$$

Then

$$\inf_{D_2 \in \mathcal{D}_2} \max_{\Delta_1 \in \mathbf{B\Delta}_1} \bar{\sigma}\left(D_2 F_u(M, \Delta_1) D_2^{-1} \right) = \frac{1}{\gamma} \tag{8.27}$$

Proof: Note that since $\mu_1(M_{11}) < 1$, the value of γ (in (8.26)) is positive. Next, let τ denote the infimum, that is

$$\tau := \inf_{D_2 \in \mathcal{D}_2} \max_{\Delta_1 \in \mathbf{B\Delta}_1} \bar{\sigma}\left(D_2 F_u(M, \Delta_1) D_2^{-1} \right) \tag{8.28}$$

We want to show that $\tau = \dfrac{1}{\gamma}$.

Let $\alpha < \gamma$. Then, from the definition of γ, there is a $\bar{D}_2 \in \mathcal{D}_2$ such that

$$\mu_{\hat{\Delta}} \begin{pmatrix} M_{11} & M_{12}\bar{D}_2^{-1} \\ \alpha \bar{D}_2 M_{21} & \alpha \bar{D}_2 M_{22} \bar{D}_2^{-1} \end{pmatrix} < 1 \tag{8.29}$$

Then Lemma 8.12 implies

$$\max_{\Delta_1 \in \mathbf{B\Delta}_1} \bar{\sigma}\left(\bar{D}_2 F_u(M, \Delta_1) \bar{D}_2^{-1}\right) < \frac{1}{\alpha}. \tag{8.30}$$

so trivially $\tau < \frac{1}{\alpha}$. This holds for any $\alpha < \gamma$, in particular for small enough $\epsilon > 0$ and $\alpha := \gamma - \epsilon$. Therefore, for $\epsilon > 0$, $\tau < \frac{1}{\gamma - \epsilon}$, and taking limits gives $\tau \leq \frac{1}{\gamma}$. Suppose it is truly less, ie. $\tau < \frac{1}{\gamma}$. Then by the definition of τ, (8.28), there is a $\breve{D}_2 \in \mathcal{D}_2$ such that

$$\max_{\Delta_1 \in \mathbf{B\Delta}_1} \bar{\sigma}\left(\breve{D}_2 F_u(M, \Delta_1) \breve{D}_2^{-1}\right) < \frac{1}{\gamma}. \tag{8.31}$$

Then Lemma 8.12 and equation (8.31) imply that

$$\mu_{\hat{\Delta}} \begin{pmatrix} M_{11} & M_{12} D_2^{-1} \\ \gamma D_2 M_{21} & \gamma D_2 M_{22} D_2^{-1} \end{pmatrix} < 1 \tag{8.32}$$

Using continuity, for small enough $\delta > 0$, we would then have

$$\mu_{\hat{\Delta}} \begin{pmatrix} M_{11} & M_{12} \breve{D}_2^{-1} \\ (\delta + \gamma) \breve{D}_2 M_{21} & (\delta + \gamma) \breve{D}_2 M_{22} \breve{D}_2^{-1} \end{pmatrix} < 1 \tag{8.33}$$

which violates the definition of γ. Hence $\tau = \frac{1}{\gamma}$ as claimed.$\sharp$

This is an interesting result. Note that the structure which we need to compute μ with respect to does not depend on $\mathbf{\Delta}_2$. If $\mu_{\hat{\mathbf{\Delta}}}$ can be computed, then, modulo the necessary search over the D_2 and α this is a useful theorem. Later, in section 10, we will use the **general optimal constant scaling theorem** to optimally scale transfer functions using **constant**, block structured scalings. This, along with the small gain theorem, will provide a method of analyzing linear, time invariant, multivariable systems with structured, time varying and/or cone bounded nonlinear perturbations.

9 A Class of Uncertain Difference Equations

9.1 Robust stability

In this section, we present some robustness results for a class of uncertain systems. The presentation centers around discrete time systems, as the explanation seems simpler, though everything done here has a continuous time analog. The required transformation for continuous time systems is discussed in the appendix.

Suppose $M \in \mathbf{C}^{(n+m)\times(n+m)}$ is given, partitioned as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \tag{9.1}$$

where $M_{11} \in \mathbf{C}^{n\times n}, M_{12} \in \mathbf{C}^{n\times m}, M_{21} \in \mathbf{C}^{m\times n}$, and $M_{22} \in \mathbf{C}^{m\times m}$. Let $\mathbf{\Delta}$ be a $m \times m$ block structure, with corresponding D scaling set denoted by $\mathcal{D}$. Suppose $\mu_{\mathbf{\Delta}}(M_{22}) < 1$. Then for every $\Delta \in \mathbf{B\Delta}$, the linear fractional transformation $F_l(M, \Delta)$ is a well defined element of $\mathbf{C}^{n\times n}$. Let $x_k \in \mathbf{C}^n$ evolve via the (possibly time varying) linear difference equation

$$x_{k+1} = F_l(M, \Delta_k)x_k \tag{9.2}$$

where for each time step k, $\Delta_k \in \mathbf{B\Delta}$. Such a system would arise if a parametrically uncertain plant, as described in section 2, had a feedback controller, that stabilized the nominal system, and we were interested in the stability of the closed loop for all the possible perturbed plants.

Consider the following three assumptions on the uncertainty Δ_k. For each k:

(a.1) $\Delta_k \in \mathbf{\Delta}$

(a.2) $\bar{\sigma}(\Delta_k) \leq 1$

(a.3) Δ_k is fixed - ie. it does not vary with k

We want to guarantee the stability of the system described in (9.2), knowing only these three assumptions.

Since **(a.3)** implies that the system is *time invariant*, the stability of the uncertain system amounts to nothing more than checking the magnitude of the eigenvalues of $F_l(M, \Delta)$ for each $\Delta \in \mathbf{B\Delta}$ and is equivalent to

$$\max_{\Delta \in \mathbf{B\Delta}} \rho(F_l(M, \Delta)) < 1$$

Recall, from section 3.1, that $\rho(\cdot)$, the spectral radius, is a special case of μ. Hence this question can be answered using Theorem 8.2, on M with an augmented structure. The augmentation is straightforward. Define

$$\tilde{\mathbf{\Delta}} = \{\text{diag}\,[\delta I_n, \Delta] : \delta \in \mathbf{C}, \Delta \in \mathbf{\Delta}\} \tag{9.3}$$

For the upper bound, which we will use later, the corresponding D scaling set will be denoted $\tilde{\mathcal{D}}$ and is of course

$$\tilde{D} = \left\{\text{diag}\,[D_1, D_2] : D_1 \in \mathbf{C}^{n \times n} \text{ is invertible}, D_2 \in \mathcal{D}\right\} \tag{9.4}$$

Theorem 9.1 *The uncertain difference equation* $x_{k+1} = F_l(M, \Delta)x_k$ *is exponentially stable for each fixed* $\Delta \in \mathbf{B\Delta}$ *if and only if*

$$\mu_{\tilde{\Delta}}(M) < 1\ , \tag{9.5}$$

where $\tilde{\mathbf{\Delta}}$ *is defined in (9.3).*

Proof: Follows by direct application of Theorem 8.2. $\sharp$

Remember, this is true for constant, but unknown Δ. If assumption **(a.3)** above is discarded, then the system is time varying. At each step, the uncertain element may be different-we only know that at each step k, it lies in the norm bounded, structured set $\mathbf{B\Delta}$. Obviously, simple spectral radius arguments do not apply. The next lemma gives a simple sufficient condition for stability.

Lemma 9.2 *If*

$$\max_{\Delta \in \mathbf{B\Delta}} \bar{\sigma}(F_l(M, \Delta)) =: \beta < 1 \tag{9.6}$$

then the uncertain, time varying difference equation (9.2) is exponentially stable, as long as Δ_k *satisfies assumptions* **(a.1)** *and* **(a.2)** *for each time step* k.

Proof: Regardless of the time variation of the perturbation, Δ_k, we get that the norm of x_k satisfies

$$\|x_k\| \le \beta^k \|x_0\| \tag{9.7}$$

which obviously decays to zero exponentially since $\beta < 1$ by assumption. $\sharp$

As stated, Lemma 9.2 is quite conservative. We can reduce the conservatism by allowing **one** state space coordinate change. The proof is simple, and is omitted.

Lemma 9.3 *If there exists an invertible $T \in \mathbf{C}^{n \times n}$ such that*

$$\max_{\Delta \in \mathbf{B\Delta}} \bar{\sigma}\left(TF_l(M,\Delta)T^{-1}\right) = \beta < 1 \tag{9.8}$$

then for each k the state x_k, of (9.2) is bounded by

$$\|x_k\| \leq \kappa(T)\,\beta^k \|x_0\| \tag{9.9}$$

where $\kappa(T)$ denotes the condition number of T.

Remark: The above reasoning is equivalent to finding *a single, quadratic Lyapunov function* for the entire set of "A" matrices

$$\{F_l(M,\Delta) : \Delta \in \mathbf{B\Delta}\}.$$

This equivalence is evident via this lemma.

Lemma 9.4 *Let $A \in \mathbf{C}^{n \times n}$ be given. There exists a Lyapunov matrix $P \in \mathbf{C}^{n \times n}$, $P = P^*, P > 0$ for $x_{k+1} = Ax_k$ if and only if there exists an invertible $T \in \mathbf{C}^{n \times n}$ such that $\bar{\sigma}(TAT^{-1}) < 1$.*

Proof: P is a Lyapunov matrix if and only if $A^*PA - P < 0$. This is equivalent to $P^{-\frac{1}{2}}A^*PAP^{-\frac{1}{2}} - I < 0$, which is the same as $\bar{\sigma}\left(P^{\frac{1}{2}}AP^{-\frac{1}{2}}\right) < 1$. ♯

Consequently, if T is a coordinate transformation that solves (9.8), then $P := T^*T$ is a single Lyapunov matrix that works. Conversely, if P is a correct Lyapunov matrix, then $P^{\frac{1}{2}}$ is a single coordinate transformation which solves (9.8).

Conceptually, the existence of a matrix T satisfying condition (9.8) can be cast as a μ test. Again we must augment the $\mathbf{\Delta}$ perturbation structure, but this time with a **full** block, since we are checking $\bar{\sigma}(\cdot)$, and we must lug around the coordinate change T. The new structure $\hat{\mathbf{\Delta}}$ is

$$\hat{\mathbf{\Delta}} = \left\{\mathrm{diag}\left[\Delta_1, \Delta_2\right] : \Delta_1 \in \mathbf{C}^{n \times n}, \Delta_2 \in \mathbf{\Delta}\right\}. \tag{9.10}$$

Now, using the Theorem 8.2, we obviously have

Theorem 9.5 *There exists an invertible $T \in \mathbf{C}^{n \times n}$ such that*

$$\max_{\substack{\Delta \in \mathbf{\Delta} \\ \bar{\sigma}(\Delta) \leq 1}} \bar{\sigma}\left(TF_l(M,\Delta)T^{-1}\right) < 1 \tag{9.11}$$

if and only if

$$\inf_{T \in \mathbf{C}^{n \times n}} \mu_{\hat{\mathbf{\Delta}}} \left[\begin{pmatrix} T & 0 \\ 0 & I_m \end{pmatrix} M \begin{pmatrix} T^{-1} & 0 \\ 0 & I_m \end{pmatrix}\right] < 1. \tag{9.12}$$

If we could calculate $\mu_{\hat{\Delta}}$ exactly, the condition (9.12) would in principle be something that could be checked, although it is unclear how the search over the T's would be done. An interesting approach we will pursue here is to substitute the $\bar{\sigma}\left(DMD^{-1}\right)$ upper bound for $\mu_{\hat{\Delta}}(\cdot)$ and see what the resulting *sufficient* condition is.

First we need the correct set $\hat{\mathcal{D}}$ for the structure $\hat{\mathbf{\Delta}}$. If $\mathcal{D}$ is the appropriate set of D scalings for $\mathbf{\Delta}$, then we define $\hat{\mathcal{D}} = \{\operatorname{diag}[d_1 I_n, D_2] : d_1 \neq 0, D_2 \in \mathcal{D}\}$. Substituting the upper bound in place of $\mu_{\hat{\Delta}}$ gives that there is a transformation T such that (9.11) is met if

$$\inf_{\substack{T \\ \hat{D}\in\hat{\mathcal{D}}}} \bar{\sigma}\left(\begin{bmatrix} d_1 I_n & 0 \\ 0 & D_2 \end{bmatrix}\begin{bmatrix} T & 0 \\ 0 & I_m \end{bmatrix} M \begin{bmatrix} T^{-1} & 0 \\ 0 & I_m \end{bmatrix}\begin{bmatrix} d_1^{-1} I_n & 0 \\ 0 & D_2^{-1} \end{bmatrix}\right) < 1. \tag{9.13}$$

The scalar d_1 is irrelevent, since it introduces no freedom that the coordinate change T didn't already provide. Absorbing d_1 into T, we rewrite (9.13) as

$$\inf_{\substack{T \\ D_2\in\mathcal{D}}} \bar{\sigma}\left(\begin{bmatrix} T & 0 \\ 0 & D_2 \end{bmatrix} M \begin{bmatrix} T^{-1} & 0 \\ 0 & D_2^{-1} \end{bmatrix}\right) < 1 \tag{9.14}$$

Note the effect the transformation T has on the minimization in (9.14). Since T is free to be any invertible matrix in $\mathbf{C}^{n\times n}$, the matrix $\operatorname{diag}[T, D_2]$ is some arbitrary element of $\tilde{\mathcal{D}}$. Hence although (9.14) is condition (9.11) with $\mu_{\hat{\Delta}}$ replaced by its upper bound, the freedom in choosing the coordinate transformation "alters" the upper bound, so that the left hand side of (9.14) is just *the* $\bar{\sigma}\left(DMD^{-1}\right)$ *upper bound for the* $\tilde{\mathbf{\Delta}}$ *structure* (not the $\hat{\mathbf{\Delta}}$ structure that was originally there in (9.11)). In other words, (9.14) is just

$$\inf_{\tilde{D}\in\tilde{\mathcal{D}}} \bar{\sigma}\left(\tilde{D}M\tilde{D}^{-1}\right) < 1, \tag{9.15}$$

and this is a sufficient condition for Theorem 9.5 to hold. We write this as a theorem.

Theorem 9.6 *If there exists a* $\tilde{D} \in \tilde{\mathcal{D}}$ *such that* $\bar{\sigma}\left(\tilde{D}M\tilde{D}^{-1}\right) = \beta < 1$, *then the uncertain, time varying, linear system*

$$x_{k+1} = F_l(M, \Delta_k)x_k \quad , \quad \Delta_k \in \mathbf{B\Delta} \tag{9.16}$$

is exponentially stable.

How do all these different conditions fit together?

9.a Theorem 9.1 showed that $\mu_{\hat{\Delta}}(M) < 1$ is both necessary and sufficient for robust stability of (9.2) **with constant, but unknown structured perturbations.**

9.b Next, Theorem 9.5 gave a necessary and sufficient condition for the existence of a single, quadratic Lyapunov function for the entire set of systems.

9.c Unfortunately, the condition in Theorem 9.5 is not really a verifiable condition, so we substituted a μ test with a $\bar{\sigma}(DMD^{-1})$ upper bound test. This gives that $\inf_{\tilde{D}\in\tilde{\mathcal{D}}} \bar{\sigma}\left(\tilde{D}M\tilde{D}^{-1}\right) < 1$ is a sufficient condition for robust stability **with unknown, time varying, structured perturbations.**

Note the similarity between the test in Theorem 9.1, and the test in Theorem 9.6. Both are associated with the $\tilde{\mathbf{\Delta}}$ structure – one involves μ and one involves the $\bar{\sigma}\left(\tilde{D}M\tilde{D}^{-1}\right)$ upper bound. Yet the conclusions each give are quite different. This sheds a little light on how fundamentally different the upper bound and μ are.

This final result described in Theorem 9.6 can also be derived from a different point of view, utilizing Lemma 8.11 from section 8.4 along with the small gain theorem.

Note that the perturbed system, (9.2), is just the loop shown below.

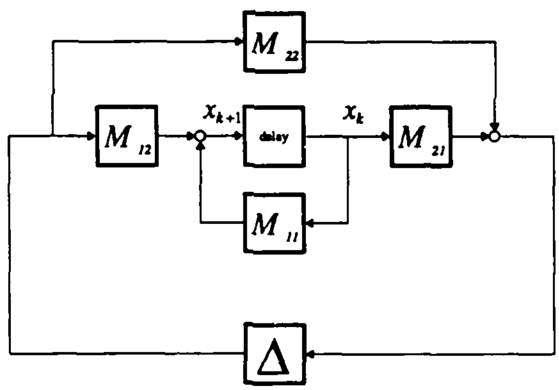


Figure 9.1 Perturbed System, Equation (9.2)

Define the transfer function $G(z) = M_{22} + M_{21}(zI - M_{11})^{-1} M_{12}$. If we can find a D_1 and D_2 (in the appropriate scaling sets, $\operatorname{diag}[D_1, D_2] \in \tilde{\mathcal{D}}$) such that

$$\bar{\sigma}\left(\tilde{D}M\tilde{D}^{-1}\right) = \bar{\sigma}\begin{pmatrix} D_1 M_{11} D_1^{-1} & D_1 M_{12} D_2^{-1} \\ D_2 M_{21} D_1^{-1} & D_2 M_{22} D_2^{-1} \end{pmatrix} < 1$$

then using the Constant D lemma, 8.11, we get

$$\|D_2 G(z) D_2^{-1}\|_\infty < 1. \tag{9.17}$$

Now for $\Delta_k \in \mathbf{\Delta}$, the two loops below are equivalent, even if Δ_k varies with k, because D_2

is **constant**, and hence it commutes with linear time varying operators as well.

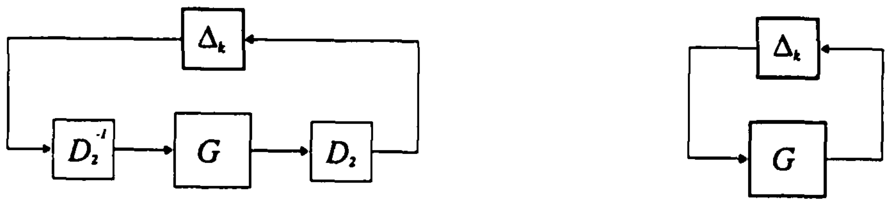


Figure 9.2 Equivalent Loops

Therefore, a trivial application of the small gain theorem along with equation (9.17) gives that the perturbed loop is stable for all varying $\bar{\sigma}(\Delta_k) \leq 1$, as expected, and in agreement with Theorem 9.6 and **9.c** above.

9.2 Robust performance

We have seen how the upper bound plays a role in determining some robustness properties of a class of uncertain difference equations when the perturbations are structured, and time varying. In this section, we continue exploring the difference between μ and the upper bound with the added objective of **performance**. Performance will be characterized in terms of the zero initial state l_2 gain from disturbance to error. Recall that for time-invariant systems, this is the same as the $\|\cdot\|_\infty$ norm of the transfer function from disturbance to error.

We begin with a matrix $M \in \mathbf{C}^{(n+n_e+m)\times(n+n_d+m)}$, partitioned obviously, and relating the variables via

$$\begin{bmatrix} x_{k+1} \\ e_k \\ z_k \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} x_k \\ d_k \\ w_k \end{bmatrix} \tag{9.18}$$

The uncertainty is "feedback" from z to w through a structured $\Delta \in \mathbf{\Delta}$, where $\mathbf{\Delta}$ is a prescribed $m \times m$ block structure. Consider a uncertain linear system (possibly time varying) driven by a disturbance input d_k, with output error e_k.

$$\begin{bmatrix} x_{k+1} \\ e_k \end{bmatrix} = F_l(M, \Delta_k) \begin{bmatrix} x_k \\ d_k \end{bmatrix} \tag{9.19}$$

With respect to this partition, $F_l(M, \Delta_k)$ is

$$\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} + \begin{bmatrix} M_{13} \\ M_{23} \end{bmatrix} \Delta_k (I - M_{33}\Delta_k)^{-1} \begin{bmatrix} M_{31} & M_{32} \end{bmatrix}$$

We need two augmented block structures. Define $\tilde{\mathbf{\Delta}}$ and $\hat{\mathbf{\Delta}}$ as

$$\tilde{\mathbf{\Delta}} := \left\{ \operatorname{diag}[\delta_1 I_n, \Delta_2] : \delta_1 \in \mathbf{C}, \Delta_2 \in \mathbf{C}^{n_d \times n_e} \right\} \tag{9.20}$$

$$\hat{\boldsymbol{\Delta}} := \left\{ \text{diag} \left[\tilde{\Delta}, \Delta \right] : \tilde{\Delta} \in \tilde{\boldsymbol{\Delta}}, \Delta \in \boldsymbol{\Delta} \right\} \tag{9.21}$$

Suppose that $\mu_{\Delta}(M_{33}) < 1$.

Theorem 9.7 (Robust Performance) *For all $\Delta \in \mathbf{B\Delta}$, the uncertain system (9.19) is stable and for zero initial state response, the error e satisfies $\|e\|_2 < \|d\|_2$ if and only if $\mu_{\hat{\Delta}}(M) < 1$.*

Proof:

$\leftarrow$ Let $\Delta \in \mathbf{B\Delta}$ be given. Since $\mu_{\hat{\Delta}}(M) < 1$, Theorem 8.2 gives

$$\mu_{\tilde{\Delta}}(F_l(M, \Delta)) < 1. \tag{9.22}$$

Stability is apparent, and the l_2 performance follows from the example in section 8.3.1.

$\rightarrow$ Essentially, the steps are reversed.

What can be concluded if the upper bound of $\mu_{\hat{\Delta}}(M)$ is less than 1?

Theorem 9.8 *Let M be given as in (9.18), along with a block structure $\boldsymbol{\Delta}$. If there is a $\hat{D} \in \hat{\mathcal{D}}$ such that*

$$\bar{\sigma}\left(\hat{D} M \hat{D}^{-1}\right) = \beta < 1$$

then for all sequences $\{\Delta_k\}_{k=0}^{\infty}$ with $\Delta_k \in \mathbf{B\Delta}$, the time varying, uncertain system

$$\begin{bmatrix} x_{k+1} \\ e_k \end{bmatrix} = F_l(M, \Delta_k) \begin{bmatrix} x_k \\ d_k \end{bmatrix} \tag{9.23}$$

is zero-input, exponentially stable, and if $x_0 = 0$, and $\{d_k\} \in l_2$, then $\|e\|_2 \leq \beta \|d\|_2$.

The results we have obtained for time varying perturbations extend to a special class of nonlinear perturbations. The appropriate definitions and assumptions are the subject of the next section.

9.3 Cone bounded nonlinearities

Let $\mathbf{N}$ be the set of nonnegative integers, and let $\mathbf{O}$ be any set.

Definition 9.9 *A unstructured, memoryless, nonlinear operator,* $S : \mathbf{N} \times \mathbf{O} \times \mathbf{C}^{n_d} \to \mathbf{C}^{n_e}$, *is cone bounded (of size* α*) if there exists a* $\alpha > 0$ *such that for all* $d \in \mathbf{C}^{n_d}$, $o \in \mathbf{O}$, *and all* $k \in \mathbf{N}$

$$\|S(k, o, d)\| \leq \alpha \|d\|.$$

In the definition, the set $\mathbf{O}$ can represent dependencies of nonlinearity S on other parameters.

Unfortunately, the notion of a $n \times n$ *repeated scalar*, cone bounded operator is trickier. The natural definition would involve a single scalar cone bounded nonlinearity, which we would then be applied separately to each of the n components of the input vector. Unfortunately, our framework cannot directly handle this, and we must treat this type of uncertainty as n **independent**, cone bounded scalar nonlinearities. So, when we refer to a cone bounded, repeated scalar block, we in fact mean a block of the form $\gamma(k, o) I_n$. Note that γ can be time varying, and depend on the other parameters which the set $\mathbf{O}$ represents. The key is that all n signals into this block get multiplied by the same scalar parameter, namely γ.

Finally, a block structured, cone bounded nonlinearity is the obvious block diagonal collection of several of these blocks. With this definition, results similar to the time varying (but linear) results are possible.

Theorem 9.10 *Let* M *be given as in (9.18), along with a block structure* $\boldsymbol{\Delta}$. *Suppose* $\Delta : \mathbf{N} \times \mathbf{O} \times \mathbf{C}^{n_e} \to \mathbf{C}^{n_d}$ *is a block structured, cone bounded nonlinearity, with cone of size 1. If there is a* $\hat{D} \in \hat{\mathcal{D}}$ *such that* $\bar{\sigma}\left(\hat{D} M \hat{D}^{-1}\right) = \beta < 1$, *then the uncertain system*

$$\begin{bmatrix} x_{k+1} \\ e_k \\ z_k \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} x_k \\ d_k \\ w_k \end{bmatrix}$$

$$w_k = \Delta(k, o_k, z_k)$$

is zero-input, exponentially stable, and if $x_0 = 0$, *and* $\{d_k\}_{k=0}^{\infty} \in l_2$, *then* $\|e\|_2 \leq \beta \, \|d\|_2$.

10 Optimal Constant D scalings for Multivariable Systems

This section combines two results from previous sections, to yield a method for sub-optimal and optimal scaling of multivariable transfer functions using constant, diagonal D matrices.

Let $G(z)$ be a given, stable, transfer function, with m inputs, and m outputs, and state space realization

$$G(z) = D + C\,(zI - A)^{-1}\,B \tag{10.1}$$

Suppose a perturbation structure $\mathbf{\Delta}_2$ is given, and is compatible with $G(z)$. That is, $\mathbf{\Delta}_2 \subset \mathbf{C}^{m\times m}$. As usual, let $\mathcal{D}_2$ denote the set of diagonal scalings (here, for simplicity, we revert back to the nonexponential notation for the D's, ie., the set $\mathcal{D}_g$ from section (5.1)) that commute with all elements of $\mathbf{\Delta}_2$.

Optimal constant scaling is the constant D scale that achieves the following infimum (if it exists, otherwise, a scaling that gets arbitrarily close)

$$\inf_{D_2\in\mathcal{D}_2} \sup_{\substack{z\in\mathbf{C}\\ |z|\geq 1}} \bar{\sigma}\left(D_2 G(z) D_2^{-1}\right) \tag{10.2}$$

Remark: This is useful because any linear perturbation, **even a time varying perturbation**, with the appropriate block diagonal structure as defined by $\mathbf{\Delta}_2$, commutes with these **constant** D scales. Therefore, for every constant $D \in \mathcal{D}$ and every operator Δ_2, with the correct block diagonal structure, the following operators are the same

$$D\Delta_2 D^{-1} = \Delta_2$$

Therefore, for any operator G, the following systems are equivalent (any solution to the loop equations in one system are also solutions to the loop equations of the other).

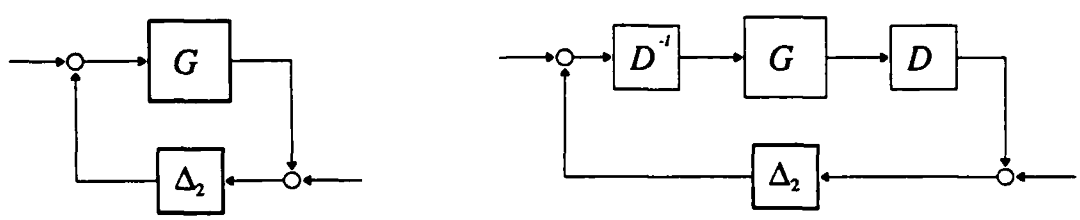


Figure 10.1 Equivalent Loops

Simple application of the small gain theorem, ([Zam] and [DesV]), on the right figure gives that if Δ_2 is a stable operator mapping $l_2 \rightarrow l_2$, and the induced norm of Δ_2, $\|\Delta_2\|$, satisfies

$$\|\Delta_2\| < \frac{1}{\|DGD^{-1}\|_\infty}$$

then the loop is stable. Hence, if we can maximize the right hand side, this will eliminate some of the conservatism in the small gain theorem due to the structure of the perturbation. This calls for a minimization of the form

$$\inf_{D \in \mathcal{D}} \|DGD^{-1}\|_\infty$$

An important point to reiterate is that the D's are constant. If they were frequency varying, then in general they would not commute with time varying Δ's, and hence the equivalence of the two figures would be invalid.

Refering back to section 8.4 we see that, at least conceptually, Theorem 8.13 gives the value of the infimum. Here we will capitalize on the additional structure that is present in this specific problem, and use the result for block structures with $f = s = 1$ which we obtained in section 7.2 to give a computationally tractable approach.

First, let $A \in \mathbf{C}^{n \times n}, B \in \mathbf{C}^{n \times m}, C \in \mathbf{C}^{m \times n}$, and $D \in \mathbf{C}^{m \times m}$ be a realization of $G(z)$. We assume G is stable, so $\rho(A) < 1$. Recall that by inverting the $\mathcal{Z}$ transform variable, we can rewrite (10.2) as

$$\inf_{D_2 \in \mathcal{D}_2} \max_{\substack{\Delta_1 \in \mathbf{\Delta}_1 \\ \bar{\sigma}(\Delta_1) \leq 1}} \bar{\sigma}\left(D_2 F_u(M, \Delta_1) D_2^{-1}\right) \tag{10.3}$$

where

$$M := \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{10.4}$$

and $\mathbf{\Delta}_1 = \{\delta I_n : \delta \in \mathbf{C}\}$.

Direct application of Theorem 8.13 from 8.4 gives,

$$\inf_{D_2 \in \mathcal{D}_2} \sup_{\substack{z \in \mathbf{C} \\ |z| \geq 1}} \bar{\sigma}\left(D_2 G(z) D_2^{-1}\right) = \frac{1}{\gamma} \tag{10.5}$$

where γ is defined by

$$\gamma := \sup_{\alpha > 0} \left\{ \alpha : \inf_{D_2 \in \mathcal{D}_2} \mu_{\hat{\Delta}} \begin{pmatrix} A & BD_2^{-1} \\ \alpha D_2 C & \alpha D_2 D D_2^{-1} \end{pmatrix} < 1 \right\} \tag{10.6}$$

using the block structure

$$\hat{\mathbf{\Delta}} := \left\{ \operatorname{diag}[\delta_1 I_n, \Delta] : \delta_1 \in \mathbf{C}, \Delta \in \mathbf{C}^{m \times m} \right\}. \tag{10.7}$$

This, $\hat{\mathbf{\Delta}}$ in (10.7), is precisely the structure we considered in section 7.2, and with respect to this structure, $\mu_{\hat{\Delta}}(M) = \inf_{\hat{D}\in\hat{\mathcal{D}}} \bar{\sigma}\left(\hat{D}M\hat{D}^{-1}\right)$. Hence the quantity γ in (10.6) can be defined in terms the upper bound, instead of μ. The expression for γ below follows immediately by substituting the infimum for μ into (10.6).

$$\gamma = \sup_{\alpha>0} \left\{\alpha : \inf_{D_2\in\mathcal{D}_2} \inf_{\substack{D_1 \\ \text{invertible}}} \bar{\sigma}\begin{pmatrix} D_1AD_1^{-1} & D_1BD_2^{-1} \\ \alpha D_2CD_1^{-1} & \alpha D_2DD_2^{-1} \end{pmatrix} < 1\right\}. \tag{10.8}$$

We state this as a theorem.

Theorem 10.1 *Let $G(z)$ and $\mathbf{\Delta}_2$ be given as in the beginning of this section. Define $\gamma \in \mathbf{R}$ by*

$$\gamma := \sup_{\alpha>0} \left\{\alpha : \inf_{\substack{D_1 \text{ invertible} \\ D_2\in\mathcal{D}_2}} \bar{\sigma}\begin{pmatrix} D_1AD_1^{-1} & D_1BD_2^{-1} \\ \alpha D_2CD_1^{-1} & \alpha D_2DD_2^{-1} \end{pmatrix} < 1\right\}. \tag{10.9}$$

Then

$$\inf_{D_2\in\mathcal{D}_2} \sup_{\substack{z\in\mathbf{C} \\ |z|\geq 1}} \bar{\sigma}\left(D_2G(z)D_2^{-1}\right) = \frac{1}{\gamma}\sharp \tag{10.10}$$

How is this computed? For a given $\alpha > 0$, we can find the infimum using the descent directions for $\bar{\sigma}$ that were presented in section 5.3. Carrying out a one dimensional search to find the correct value of γ completes the calculation.

The sufficient condition is easy, and follows directly from Lemma 8.11.

Lemma 10.2 *If there is a* $\text{diag}\,[D_1, D_2] \in \hat{\mathcal{D}}$, *and an* $\alpha > 0$ *such that*

$$\bar{\sigma}\begin{pmatrix} D_1AD_1^{-1} & D_1BD_2^{-1} \\ \alpha D_2CD_1^{-1} & \alpha D_2DD_2^{-1}, \end{pmatrix} < 1$$

then

$$\|D_2GD_2^{-1}\|_\infty \leq \frac{1}{\alpha}.$$

11 Frequency domain techniques

The most well known use of μ is as a frequency domain tool, specifically, as a generalization of the singular value tools developed in the late 70's, [DoyS]. Singular values are useful for one full block of uncertainty, but are generally conservative when the uncertainty has structure (recall, for one full block, $\mu = \bar{\sigma}$, but for other structures the gap between μ and $\bar{\sigma}$ may be arbitrarily large). Hence singular value-like frequency plots, using μ instead of $\bar{\sigma}$ can handle structured unmodeled dynamics, [DoyWS].

This section will present a simple set of modeling assumptions, along with the robustness theorems that subsequently arise. The modeling approach we adopt here is quite unsophisticated. This will help us avoid more complicated topological issues of modeling uncertainty, which would take us too far from the spirit of the research. A natural way to view uncertainty in an individual component is as follows: the only knowledge about the actual component is that it lies in some predescribed set of *possible components* (the use of μ almost **requires** that the prescribed set be defined in terms of a linear fractional transformation). Work by Vidyasagar and [FooP] has shown that the set representing the actual component should be *path connected in the graph topology.* The graph topology is a topology on the space of proper, rational transfer matrices. It was introduced in [Vid], and is best characterized in terms of coprime factorizations. We would like to bypass this issue, since it is not central to the ideas here. Moreover, obtaining necessary conditions for robust stability is much less understood in this framework. Consequently, we will be content with the simplified uncertainty modeling presented here. Fortunately, in either approach, the robustness test (using a μ framework) will still involve calculating μ on a specific nominal, closed loop transfer function.

Apart from the differences in time domains (continuous versus discrete), the results of this section are entirely equivalent to those from section 9. In effect, we replace the single μ test of Theorems 9.1 and 9.7 with a frequency varying μ test on a smaller matrix and correspondingly smaller block structure. This is possible via the maximum modulus result, Theorem 8.6. In spite of this mathematical equivalence, the results in this section are derived using a Nyquist-based argument, which is consistent with the historical development of these robustness methods.

We begin with some well known results on the stability of feedback loops.

11.1 Stability of Feedback Loops

Consider two finite dimensional, linear, time invariant systems described by the equations

$$\dot{x}_i = A_i x_i + B_i u_i$$
$$y_i = C_i x_i + D_i u_i$$

Assume that the number of outputs in system 1 equals the number of inputs to system 2, and vice versa. Hence $D_1 D_2$ is a square matrix, and we assume that $I + D_1 D_2$ is invertible. Let $M_i(s)$ denote the transfer function of each system.

Suppose that for each $i = 1, 2$, the pair (A_i, B_i) is stabilizable, and the pair (A_i, C_i) is detectable. Consider the interconnection $u_1 = y_2 + v_1$; $u_2 = v_2 - y_1$ shown below.

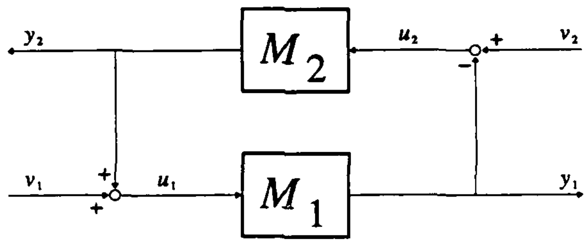


Figure 11.1 Feedback Interconnection of Two Systems

Then the internal dynamics of the interconnection, which are governed by the matrix

$$\begin{bmatrix} A_1 - B_1 (I + D_2 D_1)^{-1} D_2 C_1 & B_1 (I + D_2 D_1)^{-1} C_2 \\ -B_2 (I + D_1 D_2)^{-1} C_1 & A_2 - B_2 (I + D_1 D_2)^{-1} D_1 C_2 \end{bmatrix}$$

are stable if and only if the transfer function from $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ to $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ has all of its poles in the open left half plane (proper, rational, transfer functions with all poles in the open left half plane will be referred to as *stable*). This is easy to verify by showing that the internal dynamics are stabilizable from v, and detectable from y.

Theorem 11.1 *If both M_1 and M_2 are stable, then the interconnection is stable if and only if $(I + M_1(s)M_2(s))^{-1}$ is stable.*

Proof: All four of the transfer functions are linear combinations of I, M_1, M_2, and $(I + M_1 M_2)^{-1}$, hence, if these separately stable, all 4 of the transfer functions are. Conversely, $(I + M_1(s)M_2(s))^{-1}$ is equal to $I - H_{y_1,v_2}$, where H_{y_1,v_2} is the transfer function from v_2 to y_1. Hence $(I + M_1 M_2)^{-1}$ necessarily is stable if the interconnection is. $\sharp$

Alternatively, we have the multivariable Nyquist test, which in the case that both systems are stable, has a particularly simple form.

Theorem 11.2 *Suppose both M_1 and M_2 are stable. The interconnection is stable if and only if the Nyquist plot of* $\det(I + M_1(j\omega)M_2(j\omega))$, *does not pass through or encircle the origin. as ω varies from $-\infty \to \infty$.*

11.2 Representing unmodeled dynamics

In this section we describe a simple set of assumptions for modeling components with unmodeled dynamics. As mentioned earlier, similar, but more sophisticated assumptions exist, [FooP].

Consider a "two" input, "two" output system G described by the following state space equations

$$\begin{bmatrix} \dot{x} \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} x \\ u_1 \\ u_2 \end{bmatrix} \tag{11.1}$$

where $A \in \mathbf{R}^{n\times n}, B_i \in \mathbf{R}^{n\times n_{u_i}}, C_i \in \mathbf{R}^{n_{y_i}\times n}, D_{ij} \in \mathbf{R}^{n_{y_i}\times n_{u_i}}$. We will use this state space description to represent an uncertain component. We begin with the following assumptions:

- The nominal model for this component is given by the quadruple (A, B_2, C_2, D_{22}). The pair (A, B_2) is stabilizable, and the pair (A, C_2) is detectable.
- $\bar{\sigma}(D_{11}) \leq 1$.

The uncertainty in the component will of course be parametrized by a linear fractional transformation. Let $\mathbf{\Delta}$ be any given block structure, with overall dimensions $n_{u_1} \times n_{y_1}$. With respect to this $\mathbf{\Delta}$, define the following set of state space quadruples

$$\mathcal{R}_{\mathbf{\Delta}} = \left\{ \left(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}\right) : \tilde{A} \text{ is stable, } \tilde{D} + \tilde{C}\left(j\omega I - \tilde{A}\right)^{-1}\tilde{B} \in \mathbf{\Delta} \text{ for all } \omega \in \mathbf{R} \right\} \tag{11.2}$$

where the matrices are $\tilde{A} \in \mathbf{R}^{m\times m}, \tilde{B} \in \mathbf{R}^{m\times n_{y_1}}, \tilde{C} \in \mathbf{R}^{n_{u_1}\times m}, \tilde{D} \in \mathbf{R}^{n_{u_1}\times n_{y_1}}$ and m ranges over all nonnegative integers. Furthermore, define a subset of $\mathcal{R}_{\mathbf{\Delta}}$ as

$$\mathbf{B}\mathcal{R}_{\mathbf{\Delta}} := \left\{ \left(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}\right) \in \mathcal{R}_{\mathbf{\Delta}} : \sup_{\omega} \bar{\sigma}\left(\tilde{D} + \tilde{C}\left(j\omega I - \tilde{A}\right)^{-1}\tilde{B}\right) < 1 \right\} \tag{11.3}$$

The set of components that the pair $(G, \mathcal{R}_{\mathbf{\Delta}})$ define are

$$\begin{bmatrix} \dot{x} \\ \dot{\zeta} \\ y \end{bmatrix} = \begin{bmatrix} A + B_2\tilde{D}ZC_1 & B_1W\tilde{C} & B_2 + B_1\tilde{D}ZD_{12} \\ \tilde{B}ZC_1 & \tilde{A} + \tilde{B}ZD_{11}\tilde{C} & \tilde{B}ZD_{12} \\ C_2 + D_{21}\tilde{D}ZC_1 & D_{21}W\tilde{C} & D_{22} + D_{21}\tilde{D}ZD_{12} \end{bmatrix} \begin{bmatrix} x \\ \zeta \\ u \end{bmatrix} \tag{11.4}$$

where $Z := \left(I - D_{11}\tilde{D}\right)^{-1}$ and $W := \left(I - \tilde{D}D_{11}\right)^{-1}$, and the quadruple $\left(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}\right) \in$ $\mathbf{B}\mathcal{R}_{\mathbf{\Delta}}$.

In a diagram, this is just $F_u\left(G(s), \Delta(s)\right)$, where $\Delta(s) = \tilde{D} + \tilde{C}\left(sI - \tilde{A}\right)^{-1}\tilde{B}$. This is shown below.

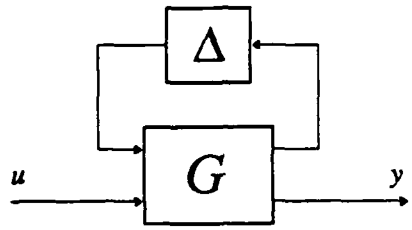


Figure 11.2 General, Uncertain Component Model

This is the general model we will use for a component with structured, unmodeled dynamics.

Remark: *To simplify the discussion, we will treat the perturbations as actual components.* This is implicit in the state-space manner that we have written the perturbed component. That way, we avoid the technical dilemma which occurs when modeling, say, the constant component $g = 1$ using a LFT on G defined as

$$G(s) := \begin{bmatrix} \frac{a}{s+a} & 0 \\ b & 1 \end{bmatrix} \tag{11.5}$$

Note that regardless of $\Delta(s)$, the linear fractional transformation $F_u(G, \Delta) = 1$, so that this does indeed represent the constant component $g = 1$. However, the μ test we are about to describe would give that the uncertainty in this component, if large enough, can cause instability in any closed loop system with this component. If we treat the uncertainty as components, then this interpretation is correct.

Finally, we define an **uncertain plant** as a linear interconnection of uncertain components, that is itself an uncertain component. Therefore, through its actual inputs and outputs, the dynamics are stabilizable and detectable. A collection of uncertain components defines a new uncertainty structure that also has the block diagonal form. Simply by reordering each of the separate uncertainties, we can assume the structure is like that defined in section 3. The plant also has a multivariable exogenous disturbance and multivariable error. These are additional injections to the component dynamics, and various internal signals from the components. Hence, the uncertain plant is described by a known dynamical system $P(s)$, and a given uncertainty set $\mathbf{\Delta}$. In particular, P is described by the state space equations

$$\begin{bmatrix} \dot{x} \\ y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} A & B_1 & B_2 & B_3 \\ C_1 & D_{11} & D_{12} & D_{13} \\ C_2 & D_{21} & D_{22} & D_{23} \\ C_3 & D_{31} & D_{32} & D_{33} \end{bmatrix} \begin{bmatrix} x \\ u_1 \\ u_2 \\ u_3 \end{bmatrix} \tag{11.6}$$

where A is stabilizable through B_3, and detectable via C_3. Let K stabilize the nominal, ie. K stabilizes the dynamic system described by A, B_3, C_3, D_{33}. The diagram below shows the perturbed plant with controller K. The signal u_3 is the manipulated variable, and this depends on the measurements, y_3, via the control law $u_3(s) = K(s)y_3(s)$. The signal u_2 is the exogenous disturbance, and y_2 is the error. A stable, finite dimensional $\Delta(s) \in \mathbf{B}\mathcal{R}_{\mathbf{\Delta}}$ is the perturbation, and this relates u_1 to y_1 via the "feedback" $u_1(s) = \Delta(s)y_1(s)$.

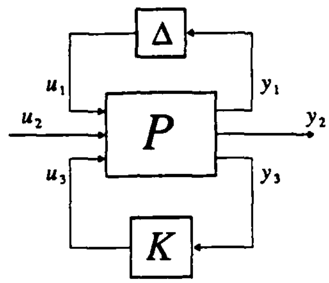


Figure 11.3 Perturbed Plant with Feedback Controller

What questions would we like to answer?

- determine whether the closed loop is stable for all stable $\Delta(s) \in \mathbf{B}\mathcal{R}_{\mathbf{\Delta}}$, and
- if so, determine how large (in $\|\cdot\|_\infty$ norm) the perturbed disturbance to error map will get.

11.3 Frequency domain robustness tests

We have the following facts/assumptions:

- The controller stabilizes the nominal, hence the internal dynamics of $F_l(P, K)$ are stable. Let $M(s) := F_l(P(s), K(s))$, the closed loop transfer function from (u_1, u_2) to (y_1, y_2). The perturbed disturbance-to-error transfer function is $F_u(M(s), \Delta(s))$.
- The perturbations are themselves viewed as stable components. Therefore, the perturbed closed loop is stable if and only if the transfer function $(I - M_{11}(s)\Delta(s))^{-1}$ is stable. As we shall now see, this can be readily cast as a μ test on the loop transfer function $M_{11}(j\omega)$.

Theorem 11.3 (Robust Stability) *The perturbed closed loop is stable for all* $\Delta(\cdot) \in \mathbf{B}\mathcal{R}_{\mathbf{\Delta}}$ *if and only if* $\sup_{\omega} \mu_{\mathbf{\Delta}}(M_{11}(j\omega)) \leq 1$.

Proof:

$\leftarrow$ As we have pointed out, we need only check the stability of the transfer function $(I - M_{11}(s)\Delta(s))^{-1}$ for each $\Delta(s) \in \mathbf{B}\mathcal{R}_{\mathbf{\Delta}}$, from Theorem 11.1. Let $\Delta(s)$ be an arbitrary element of $\mathbf{B}\mathcal{R}_{\mathbf{\Delta}}$, and suppose $\sup_{\omega} \mu_{\mathbf{\Delta}}(M_{11}(j\omega)) \leq 1$. Both are stable, therefore using Theorem 11.1, we only need to show that the Nyquist plot for $(I + M_{11}(j\omega)\Delta(j\omega))$ does not pass through or encircle the origin. For $\tilde{\epsilon} > 0$, but sufficiently small, the interconnection of M_{11} and $\tilde{\epsilon}\Delta$ will be stable by continuity of eigenvalues (or small gain theorem). Hence the nyquist plot of $(I + \tilde{\epsilon}M_{11}(j\omega)\Delta(j\omega))$ must not pass through or encircle the origin. For every $\epsilon \in [\tilde{\epsilon}, 1]$ and every $\omega \in [-\infty, \infty]$,

$$\begin{array}{c} \bar{\sigma}(\epsilon\Delta(j\omega)) < 1 \\ \mu[M_{11}(j\omega)] \leq 1 \end{array} \rightarrow \det(I + \epsilon M_{11}(j\omega)\Delta(j\omega)) \neq 0 \tag{11.7}$$

Setting $\epsilon = 1$ in (11.7) gives that the Nyquist plot for $(I + M_{11}(j\omega)\Delta(j\omega))$ **does not pass through** the origin. But, it cannot encircle the origin either. To see this, recall that for small enough ϵ, it did not encircle the origin. As $\epsilon \nearrow 1$, the Nyquist curve of $(I + \epsilon M_{11}(j\omega)\Delta(j\omega))$ deforms continuously with ϵ, and (11.7) guarantees that **it never passes through the origin**. This implies that the number of encirclements must stay the same, namely zero, so the actual perturbed loop ($\epsilon = 1$) is indeed stable. A rigorous homotopy argument for this deformation proof can be found in [CheD].

$\rightarrow$ Suppose $\sup_{\omega} \mu[M_{11}(j\omega)] > 1$. Then for some finite $\bar{\omega} \in \mathbf{R}$, $\mu[M_{11}(j\bar{\omega})] > 1$. Choose a constant, complex matrix $\Delta_c \in \mathbf{\Delta}$ such that $\det(I + M_{11}(j\bar{\omega})\Delta_c) = 0$, and $\bar{\sigma}(\Delta_c) < 1$. This is always possible. Then the interconnection with M_{11} and Δ_c has a pole at $s = j\bar{\omega}$. It is a fairly simple task [CheD] to find a $\Delta(s) \in \mathbf{B}\mathcal{R}_{\mathbf{\Delta}}$ that interpolates Δ_c at $s = j\bar{\omega}$. This choice for $\Delta(s)$ destabilizes the loop, and completes the proof. ♯

Next, we answer the question of **robust performance** - "How large does the perturbed disturbance-to-error map, $F_u(M(s), \Delta(s))$ get as Δ takes on various values in $\mathcal{R}_1\mathbf{\Delta}$?

Theorem 11.4 (Robust Performance) *Let P be an uncertain plant as defined in (11.6), $\mathbf{\Delta}$ be a given uncertainty structure, and K be a LTIFD controller that stabilizes the nominal part of P, ie. K stabilizes the quadruple (A, B_3, C_3, D_{33}). Define an augmented structure $\hat{\mathbf{\Delta}}$ as*

$$\hat{\mathbf{\Delta}} := \left\{ \mathrm{diag}[\Delta, \Delta_2] : \Delta \in \mathbf{\Delta}, \Delta_2 \in \mathbf{C}^{n_{u_2} \times n_{y_2}} \right\} \tag{11.8}$$

so that $\hat{\mathbf{\Delta}}$ *is compatible in dimension to* $M(j\omega) := F_l(P,K)(j\omega)$.

Then, the perturbed closed loop is stable, **and** $\|F_u(M,\Delta)\|_\infty \leq 1$ *for all* $\Delta(s) \in \mathbf{B}\mathcal{R}_{\mathbf{\Delta}}$ *if and only if* $\sup_\omega \mu_{\hat{\mathbf{\Delta}}}(M(j\omega)) \leq 1$.

Proof:

$\leftarrow$ First, we always have

$$\sup_\omega \mu_{\mathbf{\Delta}}(M_{11}(j\omega)) \leq \sup_\omega \mu_{\hat{\mathbf{\Delta}}}(M(j\omega)) \leq 1 \tag{11.9}$$

so using the **Robust Stability** theorem, for all such $\Delta(s)$, the perturbed loop is stable. Let $\Delta(s) \in \mathbf{B}\mathcal{R}_{\mathbf{\Delta}}$, and let ω be arbitrary. Note that $\Delta(j\omega) \in \mathbf{\Delta}$ and $\bar{\sigma}(\Delta(j\omega)) < 1$. Since $\mu_{\hat{\mathbf{\Delta}}}(M(j\omega)) \leq 1$, Theorem 8.3 implies that

$$\bar{\sigma}(F_u(M(j\omega), \Delta(j\omega))) \leq 1 \tag{11.10}$$

Therefore, for such a $\Delta(s)$, we get that $\|F_u(M,\Delta)\|_\infty \leq 1$.

$\rightarrow$ Suppose that $\sup_\omega \mu_{\hat{\mathbf{\Delta}}}(M(j\omega)) > 1$. If, in fact, $\sup_\omega \mu_{\mathbf{\Delta}}(M_{11}(j\omega)) > 1$, then the loop can be destabilized using an element of $\mathbf{B}\mathcal{R}_{\mathbf{\Delta}}$ as described in the proof of Theorem 11.3 Otherwise, choose a finite $\bar{\omega} \in \mathbf{R}$ and $\hat{\Delta} := \begin{bmatrix} \Delta_c & 0 \\ 0 & \Delta_{2_c} \end{bmatrix} \in \hat{\mathbf{\Delta}}$ such that $\bar{\sigma}\left(\hat{\Delta}\right) < 1$ and

$$\det\left(I - M(j\bar{\omega})\hat{\Delta}\right) = 0 \tag{11.11}$$

Again, use the results in [CheD] to interpolate a stable, rational $\Delta(s)$ such that $\|\Delta(s)\|_\infty < 1$ and $\Delta_c = \Delta(j\bar{\omega})$. Then,

$$\det\left(\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} - \begin{bmatrix} M_{11}(j\bar{\omega}) & M_{12}(j\bar{\omega}) \\ M_{21}(j\bar{\omega}) & M_{22}(j\bar{\omega}) \end{bmatrix}\begin{bmatrix} \Delta(j\bar{\omega}) & 0 \\ 0 & \Delta_{2_c} \end{bmatrix}\right) = 0. \tag{11.12}$$

Since $\bar{\sigma}(\Delta_{2_c}) < 1$, (11.12) implies that

$$\bar{\sigma}(F_u(M(j\bar{\omega}), \Delta(j\bar{\omega}))) > 1 \tag{11.13}$$

which proves the desired result. $\sharp$

These theorems can also be scaled so that the bound on robustness is not 1, but some other positive number. The details are the same, using the basic ideas from the theorems in section 8.2.

12 Counterexamples showing that μ need not equal the upper bound

This section shows, via two detailed examples, that $\mu(M)$ is not always equal to the $\bar{\sigma}(DMD^{-1})$ upper bound. An appealing aspect of these examples is their simplicity, each using only elementary linear algebra.

12.1 2 *repeated scalar blocks*

We begin with the block structure $s = 2$ and $f = 0$. We use the results from section 9 on uncertain difference equations to derive the counterexample.

12.1.a Let $a \in (0,1)$ and $\gamma \in (0,1)$ be given. Define the matrix $M \in \mathbf{R}^{4\times 4}$ by

$$M := \begin{bmatrix} 0 & 1 & 0 & 1 \\ \gamma & 0 & \gamma & 0 \\ 2a & 0 & a & 0 \\ 0 & -2a & 0 & -a \end{bmatrix} \tag{12.1}$$

Define a block structure $\mathbf{\Delta} := \{\delta I_{2\times 2} : \delta \in \mathbf{C}\}$. We will investigate the stability of the difference equation

$$x_{k+1} = F_l(M, \Delta)\, x_k \tag{12.2}$$

with various assumptions on the uncertainty $\Delta \in \mathbf{\Delta}$. Recall that the results of section 9 addressed just this problem.

12.1.b For all $\Delta \in \mathbf{B\Delta}$ the LFT $F_l(M, \Delta)$ is well defined, and appears as

$$F_l(M, \Delta) = \begin{bmatrix} 0 & \dfrac{1 - a\delta}{1 + a\delta} \\ \gamma \dfrac{1 + a\delta}{1 - a\delta} & 0 \end{bmatrix}. \tag{12.3}$$

Note that for each such Δ, the spectral radius of $F_l(M, \Delta)$ is simply $\sqrt{\gamma}$, which by assumption is less than 1. Therefore, **for fixed, but unknown** uncertainties, $\Delta \in \mathbf{B\Delta}$, the system in equation (12.2) is stable. Consequently, with respect to the structure $\tilde{\mathbf{\Delta}} := \{\operatorname{diag}[\delta I_{2\times 2}, \Delta] : \delta \in \mathbf{C}, \Delta \in \mathbf{\Delta}\}$, Theorem 9.1 implies that $\mu_{\tilde{\mathbf{\Delta}}}(M) < 1$.

12.1.c Consider the **time varying system**

$$x_{k+1} = F_u(M, \Delta_k)\, x_k \tag{12.4}$$

where $\Delta_k \in \mathbf{B\Delta}$ for each time step k. Take $\Delta_k := I_{2\times 2}$ when k is even, and $\Delta_k := -I_{2\times 2}$ when k is odd. Then for k even, x_{k+2} depends on x_k by the relation

$$x_{k+2} = \begin{bmatrix} 0 & \frac{1+a}{1-a} \\ \gamma\frac{1-a}{1+a} & 0 \end{bmatrix} \begin{bmatrix} 0 & \frac{1-a}{1+a} \\ \gamma\frac{1+a}{1-a} & 0 \end{bmatrix} x_k\,. \tag{12.5}$$

which simplifies to

$$x_{k+2} = \begin{bmatrix} \gamma\frac{(1+a)^2}{(1-a)^2} & 0 \\ 0 & \gamma\frac{(1-a)^2}{(1+a)} \end{bmatrix} x_k\,. \tag{12.6}$$

For any $\gamma \in (0,1)$, it is easy to chose $a \in (0,1)$ so that (12.6), and hence (12.4), is unstable. For such choices then, we must have

$$\inf_{\tilde{D}\in\tilde{\mathcal{D}}} \bar{\sigma}\left(\tilde{D}M\tilde{D}^{-1}\right) \geq 1 \tag{12.7}$$

otherwise, by **9.1.c**, the time varying system in (12.4) **would be stable** for $\Delta_k \in \mathbf{B\Delta}$, regardless of the variation with k.

Remark: A bit more analysis can show that by proper choice of γ and a, the value of

$$\inf_{\tilde{D}\in\tilde{\mathcal{D}}} \bar{\sigma}\left(\tilde{D}M\tilde{D}^{-1}\right)$$

can be made arbitrarily close to $1+\sqrt{2}$ while $\mu_{\tilde{\Delta}}(M) < 1$.

12.2 1 *repeated scalar block*, 2 *full blocks*

Next is an example for a block structure with $s = 1$ and $f = 2$. It is broken down into 8 facts.

12.2.a First, let $\mathbf{\Delta} = \{\text{diag}\,[\delta_1, \delta_2] : \delta_i \in \mathbf{C}\}$. Then (with respect to this structure) for any complex $\tau \neq 0$,

$$\mu \begin{bmatrix} 0 & \frac{1}{\tau} \\ \tau & 0 \end{bmatrix} = 1$$

This follows as a special case of Theorem 3.4.

12.2.b Let $a \in \mathbf{C}$ with $|a| < 1$. Define G on $|\delta| \leq 1$ as

$$G(\delta) = \begin{bmatrix} 0 & \frac{1+a\delta}{1-a\delta} \\ \frac{1-a\delta}{1+a\delta} & 0 \end{bmatrix} \tag{12.8}$$

Note that everywhere in the unit disk, G is defined and looks like $\begin{bmatrix} 0 & \frac{1}{\tau} \\ \tau & 0 \end{bmatrix}$. Hence from 12.2.a

$$\sup_{|\delta|\leq 1} \mu\left[G(\delta)\right] = 1 \tag{12.9}$$

12.2.c G in (12.8), is a linear fractional transformation. In particular, define the matrix M by

$$M := \begin{bmatrix} a & 0 & 2a & 0 \\ 0 & -a & 0 & -2a \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \tag{12.10}$$

It is simple to verify that for each $|\delta| \leq 1$, $G(\delta) = F_u\left(M, \delta I_{2\times 2}\right)$.

12.2.d Define $\mathbf{\Delta}_1 = \{\delta I_{2\times 2} : \delta \in \mathbf{C}\}$, and $\mathbf{\Delta}_2 = \{\operatorname{diag}\left[\delta_1, \delta_2\right] : \delta_i \in \mathbf{C}\}$. Certainly $\mu_{1,2}(M)$ makes sense (dimensions are compatible), and $\mu_{1,2}(M) \geq 1$, since $\mu_2(M_{22}) = 1$. Using (12.2.*b*) and (12.2.*c*), and Theorem 8.3, with $\beta = 1$, gives $\mu_{1,2}(M) \leq 1$. Therefore $\mu_{1,2}(M) = 1$.

12.2.e Define the correct scaling sets $\mathcal{D}_1$ and $\mathcal{D}_2$ compatible with $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$. For any $\beta \geq 1$, and any $D_2 \in \mathcal{D}_2$

$$\sup_{|\delta|\leq \frac{1}{\beta}} \bar{\sigma}\left(D_2 F_u\left(M, \delta I_{2\times 2}\right) D_2^{-1}\right) \geq \frac{\beta + |a|}{\beta - |a|}. \tag{12.11}$$

This follows from the fact that for any $D_2 \in \mathcal{D}_2$,

$$D_2 F_u\left(M, \delta I_{2\times 2}\right) D_2^{-1} = \begin{bmatrix} 0 & \dfrac{d_1}{d_2}\dfrac{1+a\delta}{1-a\delta} \\ \dfrac{d_2}{d_1}\dfrac{1-a\delta}{1+a\delta} & 0 \end{bmatrix} \tag{12.12}$$

and from the behavior of the nonzero elements of $G(\delta)$ on the edge of disks of radius $\frac{1}{\beta}$, which is shown in the figure below.

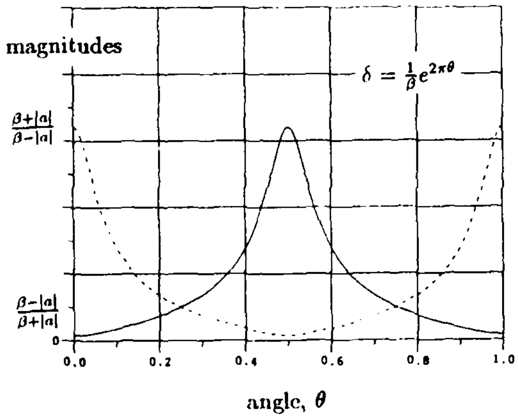


Figure 12.1 Magnitudes of Nonzero Elements of $G(\delta)$

12.2.f Fact: Let $\gamma > 0$. If there is a $\Delta_1 \in \mathbf{\Delta}_1$, $\bar{\sigma}(\Delta_1) \leq \frac{1}{\gamma}$ *such that*

- $I - M_{11}\Delta_1$ is invertible
- $\bar{\sigma}\left[F_u(M, \Delta_1)\right] \geq \gamma$

then

$$\inf_{D_1 \in \mathcal{D}_1} \bar{\sigma}\left[\begin{pmatrix} D_1 & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}\begin{pmatrix} D_1^{-1} & 0 \\ 0 & I \end{pmatrix}\right] \geq \gamma. \tag{12.13}$$

This fact is simply the contrapositive of Lemma 8.11.

12.2.g If we choose a $\beta \geq 1$ such that $\frac{\beta+|a|}{\beta-|a|} \geq \beta$, then we can apply the results from (12.2.e) and (12.2.f) above to conclude that

$$\inf_{D_1, D_2} \bar{\sigma}\left[\begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}\begin{pmatrix} D_1^{-1} & 0 \\ 0 & D_2^{-1} \end{pmatrix}\right] \geq \beta. \tag{12.14}$$

The logic is as follows: first suppose β is chosen so that $\frac{\beta+|a|}{\beta-|a|} \geq \beta$. Then from equation (12.12) we know that for every $D_2 \in \mathcal{D}_2$, there is a $\delta \in \mathbf{C}$ with $|\delta| \leq \frac{1}{\beta}$ such that

$$\bar{\sigma}\left(D_2 F_u\left(M, \delta I_{2\times 2}\right) D_2^{-1}\right) \geq \beta \tag{12.15}$$

This satisfies the conditions of (12.2.f), therefore, for each $D_2 \in \mathcal{D}_2$

$$\inf_{D_1 \in \mathcal{D}_1} \bar{\sigma}\left[\begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}\begin{pmatrix} D_1^{-1} & 0 \\ 0 & D_2^{-1} \end{pmatrix}\right] \geq \beta. \tag{12.16}$$

Carrying out the infimum over $\mathcal{D}_2$ in (12.16) yields

$$\inf_{D\in\mathcal{D}} \bar{\sigma}\left(DMD^{-1}\right) \geq \beta \tag{12.17}$$

Therefore the question becomes: "What is that largest β such that $\frac{\beta+|a|}{\beta-|a|} \geq \beta$?" Simple algebra gives that this largest β is $\beta = \frac{|a|+1+\sqrt{|a|^2+6|a|+1}}{2}$. Note that as $|a| \nearrow 1$, the quantity $\beta \nearrow 1+\sqrt{2}$.

12.2.h In summary: Let $\epsilon > 0$. Choose $a \in \mathbf{C}, |a| < 1$ such that

$$\frac{|a|+1+\sqrt{|a|^2+6|a|+1}}{2} > 1+\sqrt{2}-\epsilon. \tag{12.18}$$

Define M as in (12.10). Then, with respect to the augmented structure described in (12.2.d), $\mu(M) = 1$ but $\inf_{D\in\mathcal{D}} \bar{\sigma}\left(DMD^{-1}\right) > 1+\sqrt{2}-\epsilon$.

This example eliminates many other block structures as well. Since the *full* blocks were 1×1 in this example, they may be viewed as *repeated scalar* blocks instead. Therefore, this counterexample works for $s=2, f=1$, and $s=3, f=0$ too.

12.3 Conclusions

In light of this example, it appears that the upper bound can be quite far from the actual value of μ, especially when $s \neq 0$. For instance, in this example, the upper bound (in the limit) equals $\left(1+\sqrt{2}\right) \times \mu$. Limited computing experience with uncertainty structures having $s \neq 0$ indicates that there is often a gap, though usually not as large. For block structures with no repeated scalar blocks, $s = 0$, this contrasts directly with our computational experience. In that case, the worst known ratio of upper bound to μ is 1.14, [MorD], and usually, it is much closer to 1. Given that the upper bound can be computed, and in general, it is impossible to verify that a lower bound is indeed μ, how should this all be interpreted?

Suppose an uncertainty structure has only full blocks, and the perturbations are modeled as linear, time invariant. Using the constant, state space μ test in [DoyP] requires that the actual uncertainty structure be augmented with a large (size of state dimension) repeated scalar block. In view of the counterexample, it is likely that the upper bound will not equal μ, and the conclusions will be conservative. In this situation, a frequency domain upper bound test, [DoyWS], is appropriate, since it scales (a peak > 1 **does** give useful information), and with this block structure, we always have found μ and the upper bound very close. It is important to realize that the frequency domain test only gives conclusions about linear, time invariant perturbations.

If the perturbations are time varying and/or nonlinear, then, in general the frequency domain tests are not valid, though [Saf2] derives conditions on the frequency dependent scalings which allow for conclusions about slope bounded nonlinearities. The upper bound approaches based on **constant matrix operations** (for example, the optimal constant scaling, section 10), handle this type of uncertainty, and the motivation which led to their development was the relationship between μ and the upper bound, and the role this difference plays in the behavior of linear fractional transformations.

13 A power method for the structured singular value

This section presents an iterative algorithm to compute lower bounds for the structured singular value. The algorithm resembles a mixture of power methods for eigenvalues and singular values, which is not surprizing, since the structured singular value can be viewed as a generalization of both. If the algorithm converges, a lower bound for μ results. We prove that μ is always an equilibrium point of the algorithm, however, since in general there are many equilibrium points, we also discuss heuristic ideas to achieve convergence.

In [FanT], the calculation of μ is reformulated as a smooth optimization problem. As with all of the known exact expressions for μ, the function that is to be maximized has local maximum which are not global, so in general the method yields only lower bounds for μ. Similar comments can be made for the ideas in [Doy] and [Hel], as well as the algorithm in this paper. The contribution here is yet another lower bound algorithm to aid in the analysis of robustness of systems with structured uncertainty. This section addresses the lower bound, and develops a power algorithm aimed at quickly finding local maximums of $r:\mathbf{B\Delta}\to\mathbf{R}$, defined by $r(\Delta)=\rho(\Delta M)$. Some of the results are generalizations of those found in [DanKL].

Since we will be interested in local maximums of the function $r(\Delta)=\rho(\Delta M)$, we begin with some facts from perturbation theory, which will assist in characterizing local phenomena.

13.1 Matrix Facts

13.1.1 Derivatives of eigenvalues

In this section we review the differentiablity properties of eigenvalues and eigenvectors of matrices depending analytically on a real variable. All material comes from [Kat].

Suppose $M:\mathbf{R}\to\mathbf{C}^{n\times n}$ is an analytic function of the real parameter t. If λ_o is a eigenvalue of $M_o:=M(0)$ *of multiplicity one*, then for some open interval containing 0, this eigenvalue is a analytic function of t, as are the eigenvectors associated with it. That is, suppose there are nonzero $x_o, y_o\in\mathbf{C}^n$, satisfying

$$\begin{aligned} y_o^* x_o &= 1 \\ M_o x_o &= \lambda_o x_o \\ M_o^* y_o &= \lambda_o y_o \end{aligned} \tag{13.1}$$

Then there is an $\epsilon>0$ and analytic functions $x:(-\epsilon,\epsilon)\to\mathbf{C}^n$, $y:(-\epsilon,\epsilon)\to\mathbf{C}^n$, $\lambda:(-\epsilon,\epsilon)\to$

$\mathbf{C}$ such that for all $t \in (-\epsilon, \epsilon)$

$$\begin{aligned} y^* x &= 1 \\ Mx &= \lambda x \\ M^* y &= \lambda y \end{aligned} \tag{13.2}$$

This follows from [Kat]. Hence, we can differentiate and obtain

$$\dot{\lambda}(0) = y_o^* \dot{M}(0) x_o \tag{13.3}$$

13.1.2 Linear algebra lemmas

The next two lemmas are elementary linear algebra. They will be used in the main theorem of the next section.

Lemma 13.1 *Let $y, x \in \mathbf{C}^n$ with $y \neq 0$ and $x \neq 0$. There exists $d \in \mathbf{R}, d > 0$, such that $y = dx$ if and only if $Re\,(y^* W x) \leq 0$ for every $W \in \mathbf{C}^{n \times n}$ satisfying $W + W^* \leq 0$.*

Proof: The "only if" is obvious, so we just prove the "if". As usual, let y_i and x_i denote the i'th element of y and x, and $W_{i,j}$ denote the i,j element of $W \in \mathbf{C}^{n \times n}$. Begin by letting W be zero everywhere, except in the i,i element, and set $W_{i,i} = \sigma_i + j\omega_i$ for some $\sigma_i \leq 0$ and $\omega_i \in \mathbf{R}$. Obviously W satisfies the hypothesis. Then

$$Re\,(y^* W x) = \sigma_i \; Re\,(\bar{y}_i x_i) - \omega_i \; Im\,(\bar{y}_i x_i)$$

If $Im\,(\bar{y}_i x_i) \neq 0$, then it would be possible to choose $\omega_i \in \mathbf{R}$ to violate the $Re\,(y^* W x) \leq 0$ hypothesis. Hence $Im\,(\bar{y}_i x_i) = 0$. Similarly, with the only restriction on σ_i being $\sigma_i \leq 0$, we must have $Re\,(\bar{y}_i x_i) \geq 0$. Therefore, for each i, we can write

$$\begin{aligned} y_i &= s_i e^{j\theta_i} \\ x_i &= r_i e^{j\psi_i} \end{aligned}$$

where $s_i \geq 0, r_i \geq 0$, and $\theta_i, \psi_i \in \mathbf{R}$. From the above discussion, it is clear that for each i,

$$s_i = 0 \quad \text{or} \quad r_i = 0 \quad \text{or} \quad \theta_i = \psi_i. \tag{13.4}$$

Now, let $l \neq k$ be two integers $\leq n$. Let $\omega \in \mathbf{R}$ be arbitrary. Define a matrix W by $W_{l,k} := -e^{-j\omega}, W_{k,l} := e^{j\omega}$ and zero everywhere else. Note that $W + W^* = 0$, so trivially W satisfies the hypothesis. In this case

$$Re\,(y^* W x) = -s_l r_k \cos(\theta_k - \psi_l - \omega) + s_k r_l \cos(\theta_l - \psi_k + \omega)$$

Since ω is free and neither x or y is 0, we have for all i, $s_i = 0$ if and only if $r_i = 0$. Consequently, suppose that $s_k \neq 0$ and $s_l \neq 0$. Recall from (13.4) that this means

$\psi_k = \theta_k$ and $\psi_l = \theta_l$. We claim that $s_l r_k = s_k r_l$. To see this, suppose instead that $s_k r_l \neq s_l r_k$. By choosing $\omega := \theta_k - \theta_l$ or $\omega := \pi + \theta_k - \theta_l$, we get that

$$Re\,(y^* W x) = |s_k r_l - s_l r_k| > 0$$

This contradicts the original assumptions, hence we must have $s_l r_k = s_k r_l$. Therefore

$$\frac{s_l}{r_l} = \frac{s_k}{r_k}$$

for every $k \neq l$ with $s_k \neq 0$ and $s_l \neq 0$. Define $d > 0$ to be this ratio. For every i, we have $y_i = dx_i$ as desired. ♯

Lemma 13.2 *Let a and b be two nonzero vectors in $\mathbf{C}^n$. Then there exists a hermitian, positive definite $D \in \mathbf{C}^{n \times n}$, such that $Db = a$ if and only if $b^* a \in (0, \infty)$.*

Proof: Again, the "only if" is easy. Conversely, suppose that $\|b\| = 1$. If not, simply scale appropriately. Let $B_\perp \in \mathbf{C}^{n \times (n-1)}$ such that the matrix $K := [b\, B_\perp] \in \mathbf{C}^{n \times n}$ is unitary. Decompose a in this basis, ie. find a scalar $\alpha \in \mathbf{C}$ and $\zeta \in \mathbf{C}^{n-1}$ such that

$$a = \alpha b + B_\perp \zeta$$

By assumption, α is real and positive. Let $W \in \mathbf{C}^{(n-1)\times(n-1)}$ be any hermitian matrix such that $W - \frac{1}{\alpha}\zeta\zeta^*$ is positive definite. It is simple to check that

$$D := K \begin{bmatrix} \alpha & \zeta^* \\ \zeta & W \end{bmatrix} K^*$$

works. ♯

13.2 Decomposition at μ

We need to define a set $\mathcal{D}_{sd}$, similar to $\mathcal{D}$ from section 3.1. It is the same as $\mathcal{D}$, except the elements are restricted only to be positive semi-definite, rather than positive definite.

$$\mathcal{D}_{sd} = \left\{ \text{diag}\ \left[D_1, \ldots, D_s, d_1 I_{m_1}, \ldots, d_f I_{m_f}\right] : D_i \in \mathbf{C}^{r_i \times r_i}, D_i = D_i^* \geq 0, d_i \in \mathbf{R}, d_i \geq 0 \right\} \tag{13.5}$$

Theorem 13.3 *Let $M \in \mathbf{C}^{n \times n}$ be given, and suppose $\lambda_o > 0$ is a distinct eigenvalue of M, with right and left eigenvectors x and y respectively, and $y^* x = 1$. Suppose that $\rho(M) = \lambda_o$. If the function $r : \mathbf{B\Delta} \to \mathbf{R}$ defined by $r(\Delta) = \rho(\Delta M)$ has a local maximum (with respect to $\mathbf{B\Delta}$) at $\Delta = I$ then there exists a $D \in \mathcal{D}_{sd}$ such that $y = D^2 x$.*

Proof: Let $G \in \mathbf{\Delta}$ with $G + G^* \leq 0$. Obviously, G appears as

$$\mathrm{diag}\,[g_1 I_{r_1}, \ldots, g_s I_{r_s}, G_1, \ldots, G_f] \tag{13.6}$$

where $Re(g_i) \leq 0$, and $G_j + G_j^* \leq 0$ for all i and j. Obviously, at $t = 0$, $e^{Gt} = I$, and $e^{Gt} \in \mathbf{B\Delta}$ for all $t \geq 0$. Define a matrix function $W : \mathbf{R} \rightarrow \mathbf{C}^{n \times n}$ by $W(t) := e^{Gt}M$. Note that at $t = 0$, λ_o is a simple eigenvalue of $W(0)$, with x and y the right and left eigenvectors. For some nonempty interval containing 0, this eigenvalue is always simple, and hence there is an analytic function of the real variable t, $\lambda(t)$, defined on that interval, such that $\lambda(t)$ is and eigenvalue of $W(t)$ for all t and $\lambda(0) = \lambda_o$. It is easy to calculate $\dot{\lambda}(0)$, namely

$$\dot{\lambda}(0) = y^* \dot{W}(0) x = \lambda_o y^* G x \tag{13.7}$$

By hypothesis, $\lambda_o > 0, \rho(M) = \lambda_o$ and the function $\rho(\Delta M)$ has a local maximum at $\Delta = I$. Therefore

$$\mathrm{Re}\left(\frac{d}{dt}\lambda(t)\bigg|_{t=0} \right) \leq 0 \tag{13.8}$$

which says that the magnitude of λ must be nonincreasing at $t = 0$. Partition x and y compatibly with the block structure $\mathbf{\Delta}$,

$$x = \begin{bmatrix} x_{r_1} \\ x_{r_2} \\ \vdots \\ x_{r_s} \\ x_{m_1} \\ x_{m_2} \\ \vdots \\ x_{m_f} \end{bmatrix}, \quad y = \begin{bmatrix} y_{r_1} \\ y_{r_2} \\ \vdots \\ y_{r_s} \\ y_{m_1} \\ y_{m_2} \\ \vdots \\ y_{m_f} \end{bmatrix} \tag{13.9}$$

where $x_{r_i}, y_{r_i} \in \mathbf{C}^{r_i}$ and $x_{m_j}, y_{m_j} \in \mathbf{C}^{m_j}$ for each i and j. Using this "block notation", and substituting (13.6) and (13.7) into (13.8) yields

$$Re\left(\sum_{i=1}^{s} g_i y_{r_i}^* x_{r_i} \;+\; \sum_{j=1}^{f} y_{m_j}^* G_j x_{m_j} \right) \leq 0. \tag{13.10}$$

This must hold for *arbitrary* $G \in \mathbf{\Delta}$ *satisfying* $G + G^* \leq 0$. Applying lemmas 3.1 and 3.2, we conclude that for each i, there is a $D_i = D_i^* \in \mathbf{C}^{n \times n}, D_i \geq 0$ such that $y_{r_i} = D_i x_{r_i}$ and for each j, there is a $d_j \in \mathbf{R}, d_j \geq 0$ such that $y_{m_j} = d_j x_{m_j}$. Arranging all of these D_i's and d_j's into one block diagonal D, and taking the hermitian square root proves the lemma. $\sharp$

Remark: The only restrictive assumption we have made in the above lemma is that the eigenvalue λ_o is distinct. This assures differentiabilty. Since λ_o is a solution of a $\max\limits_{\Delta \in \mathbf{B\Delta}} \max\limits_{i} |\lambda_i(M\Delta)|$, it is likely that at the maximum it will be distinct.

13.3 Decomposition

Recall the definition of ∇ from section 7. It was introduced to find descent directions for $\bar{\sigma}$. We will generalize the definition to be valid for any singular value, not just $\bar{\sigma}$.

Let M be a complex matrix with SVD

$$M = \beta UV^* + U_2\Sigma_2 V_2^*. \tag{13.11}$$

In this setting, β is any singular value of M, not necessarily $\bar{\sigma}(M)$, but none of the singular values in Σ_2 should equal β. We use the integer $r > 0$, to denote the multiplicity of β. Hence $U, V \in \mathbf{C}^{n\times r}, U^*U = V^*V = I_r, U_2, V_2 \in \mathbf{C}^{n\times(n-r)}, U_2^*U_2 = V_2^*V_2 = I_{n-r}$.

We proceed to define the set $\nabla_{M,\beta}$. Partition U and V compatibly with $\mathbf{\Delta}$ as

$$U = \begin{bmatrix} A_1 \\ \vdots \\ A_s \\ E_1 \\ \vdots \\ E_f \end{bmatrix} \qquad V = \begin{bmatrix} B_1 \\ \vdots \\ B_s \\ F_1 \\ \vdots \\ F_f \end{bmatrix} \tag{13.12}$$

where $A_i, B_i \in \mathbf{C}^{r_i\times r},\ E_i, F_i \in \mathbf{C}^{m_i\times r}$.

For $\eta \in \mathbf{C}^r$, with $\|\eta\| = 1$, define the following components

$$\begin{aligned} P_i^\eta &= A_i\eta\eta^*A_i^* - B_i\eta\eta^*B_i^* \\ p_j^\eta &= \eta^*\left(E_j^*E_j - F_j^*F_j\right)\eta. \end{aligned} \tag{13.13}$$

Let $\nabla_{M,\beta} \subset \mathbf{X}$ be the set of all such P^η.

$$\nabla_{M,\beta} := \left\{ \text{diag } \left[P_1^\eta, \ldots, P_s^\eta, p_1^\eta, \ldots, p_{f-1}^\eta\right] : P_i^\eta, p_i^\eta \text{ in } (13.13), \eta \in \mathbf{C}^r, \|\eta\| = 1 \right\}. \tag{13.14}$$

Note that here we use two subscripts on ∇. The first is the matrix, and the second is the singular value in question. The main reason we introduce $\nabla_{M,\beta}$ here is that if there is a singular value, β, of M, and $0 \in \nabla_{M,\beta}$, then β is a lower bound for $\mu(M)$.

Theorem 13.4 *Let M and a compatible block structure $\mathbf{\Delta}$ be given. Suppose β is a singular value of M with multiplicity r. Define $\nabla_{M,\beta}$ as in (13.14). Then $0 \in \nabla_{M,\beta}$ if and only if there exists a vector $x \in \mathbf{C}^n$, a matrix $X_\perp \in \mathbf{C}^{n\times n}$, a matrix $Q \in \mathcal{Q}$, such that $\|x\| = 1$, $x^*X_\perp = 0$, $X_\perp x = 0$, and*

$$QM = \beta xx^* + X_\perp \tag{13.15}$$

Proof: Let the SVD of M be

$$M = \beta UV^* + U_2\Sigma_2 V_2^* \tag{13.16}$$

$\rightarrow$ If $0 \in \nabla_{M,\beta}$ then there exists a $\eta \in \mathbf{C}^r$, $\|\eta\| = 1$ such that

$$\begin{array}{ll} A_i\eta\eta^* A_i{}^* - B_i\eta\eta^* B_i{}^* = 0 & i \leq s \\ \eta^* \left(E_j{}^* E_j - F_j{}^* F_j\right)\eta = 0 & j \leq f-1 \end{array}$$

These relations, and the partition in (13.12) imply that there is $Q \in \mathcal{Q}$ such that

$$QU\eta = V\eta \tag{13.17}$$

Define $x \in \mathbf{C}^n$ as the above: $x := QU\eta = V\eta$. Since $\|\eta\| = 1$ and U and V are isometries, $\|x\| = 1$. Simple manipulation of (13.16) and (13.17) gives

$$(QM)\,x = (QM)\,V\eta = \beta QU\eta = \sigma V\eta = \beta x$$

$$x^*\,(QM) = \eta^* U^* Q^*\,(QM) = \beta\eta^* V^* = \beta x^*$$

Defining $X_\perp := M - \beta xx^*$ completes the decomposition

$\leftarrow$ Suppose Q, x, and $X_\perp$ are given as in the hypothesis, so that

$$QM = \beta xx^* + X_\perp$$

Define $\tilde{M} := QM$. A singular value decomposition of $\tilde{M}$ is

$$\tilde{M} = \beta\,(QU)\,(V)^* + (QU_2)\,\Sigma_2\,(V_2)^*$$

Hence β is a singular value of $\tilde{M}$, and $\tilde{M}x = \beta x$ and $M^*x = \beta x$, and so there exists a vector $\eta \in \mathbf{C}^r$, $\|\eta\| = 1$ such that

$$x = QU\eta = V\eta$$

This implies that $0 \in \nabla_{M,\beta}$ as desired. $\sharp$

It is obvious from the decomposition that β is a lower bound for $\mu\,(M)$ since β is an **eigenvalue** of $\tilde{M} = QM$. The following corollary follows immediately.

Corollary 13.5 *Let M and a compatible block structure $\mathbf{\Delta}$ be given. Suppose $D \in \mathcal{D}$, and that β is a singular value of DMD^{-1} with multiplicity r. Define $\nabla_{DMD^{-1},\beta}$ as above. Then $0 \in \nabla_{DMD^{-1},\beta}$ if and only if there exists a vector $x \in \mathbf{C}^n$, a matrix $X_\perp \in \mathbf{C}^{n\times n}$, and a matrix $Q \in \mathcal{Q}$ such that $\|x\| = 1$, $x^*X_\perp = 0$, $X_\perp x = 0$, and*

$$QDMD^{-1} = \beta xx^* + X_\perp \tag{13.18}$$

The main result of this section is that there is (almost) always a decomposition as in (13.18) with $\beta = \mu(M)$ (remember, any β satisfying (13.18) is a lower bound for $\mu(M)$). A preliminary result toward that result is next.

Theorem 13.6 *Let $Q_o \in \mathcal{Q}$ be the optimizer for $\max_{Q\in\mathcal{Q}}(QM)$, and suppose that the eigenvalue associated with $\rho(Q_oM)$ is distinct, call it λ_o, and λ_o is real and positive. If x and y are the right and left eigenvectors of the eigenvalue λ_o, then there exists a $D \in \mathcal{D}_{sd}$ such that*

$$\begin{aligned} Q_oMx &= \lambda_o x \\ x^*D^2Q_oM &= \lambda_o x^* D^2 \end{aligned} \tag{13.19}$$

Remark: If we consider local maximums of a function $\tilde{r}: \mathcal{Q} \to \mathbf{R}$ given by $\tilde{r}(Q) = \rho(QM)$, then the above theorem is not true. For $\tilde{r}$ as defined here, there exist examples where $\tilde{r}$ has a local maximum, but the decomposition described in (13.19) does not exist.

Proof: By Theorem 13.3, any maximizer of $\max_{Q\in\mathcal{Q}}(QM)$, is also a maximizer of $\max_{Q\in\mathbf{B}\Delta}(\Delta M)$. Define $\tilde{M} := Q_oM$, then $\Delta = I$ is a local maximizer for $\max_{Q\in\mathbf{B}\Delta}\left(\Delta\tilde{M}\right)$. Apply lemmas 13.1 and 13.2 to prove the theorem. ♯

In order to state the main theorem, we introduce some additional notation: partition the vectors x and y compatibly with the block structure,

$$x = \begin{bmatrix} x_{r_1} \\ x_{r_2} \\ \vdots \\ x_{r_s} \\ x_{m_1} \\ x_{m_2} \\ \vdots \\ x_{m_f} \end{bmatrix}, \quad y = \begin{bmatrix} y_{r_1} \\ y_{r_2} \\ \vdots \\ y_{r_s} \\ y_{m_1} \\ y_{m_2} \\ \vdots \\ y_{m_f} \end{bmatrix} \tag{13.20}$$

with $x_{r_i}, y_{r_i} \in \mathbf{C}^{r_i}$ and $x_{m_j}, y_{m_j} \in \mathbf{C}^{m_j}$. We call these the "block components of x and y".

Theorem 13.7 *Let the assumptions of Theorem 13.6 hold. Consider the block components of the eigenvectors x and y as in (13.20). If for all i, $y_{r_i}^* x_{r_i} \neq 0$, and for all j, neither x_{m_j} nor y_{m_j} are the 0 vector, then there exists a $D \in \mathcal{D}$ such that*

$$\begin{aligned} Q_oDMD^{-1}(Dx) &= \lambda_o(Dx) \\ (Dx)^* Q_oDMD^{-1} &= \lambda_o(Dx)^*. \end{aligned} \tag{13.21}$$

Remark: This result was first shown in [FanT], for the case of $s = 0$.

Proof: These additional assumptions guarantee that the D's in Theorem 13.6 are in fact positive definite, rearranging equation (13.19) gives (13.21). **This is** the decomposition, since Dx is both a right and left eigenvector of $Q_o DMD^{-1}$ associated with the eigenvalue λ_o. $\sharp$

13.4 Lower bound power algorithm

How can this decomposition be used? In this section, we propose an iterative algorithm to find such decompositions, and therefore get lower bounds for μ. The possible advantage this algorithm has over finding local maximums of the max $\rho(QM)$ lower bound is that there will be no costly eigenvalue/eigenvector evaluations, which would be necessary for cost/gradient calculations. Numerical experimentation indicates that the algorithm often completes successfully and quickly.

Rewriting (13.21), we want to find a $Q \in \mathcal{Q}, D \in \mathcal{D}_g$, $\beta > 0$, and $x \in \mathbf{C}^n$ with $\|x\| = 1$ such that

$$\begin{aligned} QDMD^{-1}x &= \beta x \\ D^{-1}M^*DQ^*x &= \beta x \end{aligned}$$

which can be rewritten as

$$\begin{aligned} M\left(D^{-1}x\right) &= \beta\left(D^{-1}Q^*x\right) \\ M^*\left(DQ^*x\right) &= \beta\left(Dx\right). \end{aligned}$$

For a given D, Q, and x, define vectors a, b, z, and w by

$$\begin{aligned} b &:= D^{-1}x \\ a &:= D^{-1}Q^*x \\ z &:= DQ^*x \\ w &:= Dx \end{aligned} \tag{13.22}$$

With this definition, we have $Mb = \sigma a$ and $M^*z = \sigma w$. We can eliminate x from (13.22), and redefine $D = D^2$ to get

$$\begin{aligned} b &= Qa \\ z &= Q^*w \\ z &= Da \\ b &= D^{-1}w \end{aligned}$$

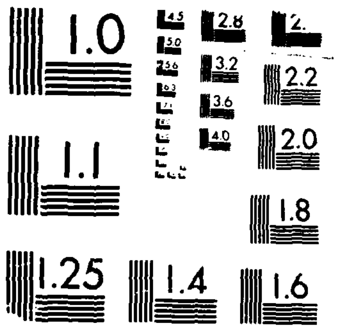
We would like to write these four new relationships in a manner that does not involve the matrices Q and D. With a few technical conditions, this can be done. In order to simplify the upcoming formulas, we will consider a block structure with $s = 1, f = 1$. *By simply duplicating the appropriate formulas for additional blocks, it is straightforward to extend the algorithm to more general structures.* Hence the sets $\mathcal{D}$ and $\mathcal{Q}$ look like

$$\mathcal{D} = \left\{\operatorname{diag}\left[D_1, d_1 I_{m_1}\right] : D_1 \in \mathbf{C}^{r_1 \times r_1}, D_1 = D_1^*, d_1 \in \mathbf{R}\right\} \tag{13.23}$$

$$\mathcal{Q} = \left\{\operatorname{diag}\left[q_1 I_{r_1}, Q_2\right] : q_1^* q_1 = 1, Q_2 \in \mathbf{C}^{m_1 \times m_1}, Q_2^* Q_2 = I\right\}. \tag{13.24}$$

AD-A194 250 ROBUST CONTROL OF MULTIVARIABLE AND LARGE SCALE SYSTEMS 2/2
(U) HONEYWELL SYSTEMS AND RESEARCH CENTER MINNEAPOLIS
MN A PACKARD ET AL. 23 MAR 88 88SRC11 AFOSR-TR-88-0450

UNCLASSIFIED F49620-86-C-0001 F/G 12/3 NL



MICROCOPY RESOLUTION TEST CHART
IREAU STANDARDS-1963-A

With respect to this, we will partition the vectors accordingly, so $z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$, where $z_1 \in \mathbf{C}^{r_1}$ and $z_2 \in \mathbf{C}^{m_1}$, and likewise for the other vectors.

Lemma 13.8 *Let r_1 and m_1 be positive integers. Let $z_1, w_1, b_1, a_1 \in \mathbf{C}^{r_1}$ and $z_2, w_2, b_2, a_2 \in \mathbf{C}^{m_1}$ be nonzero vectors with $a_1^* w_1 \neq 0$. Then, there exists a $D \in \mathcal{D}$, and $Q \in \mathcal{Q}$ such that*

$$\begin{aligned} b &= Qa \\ z &= Q^* w \\ z &= Da \\ b &= D^{-1} w \end{aligned}$$

if and only if

$$\begin{aligned} z_1 &= \frac{a_1^* w_1}{|a_1^* w_1|} w_1 \\ z_2 &= \frac{\|w_2\|}{\|a_2\|} a_2 \\ b_1 &= \frac{w_1^* a_1}{|w_1^* a_1|} a_1 \\ b_2 &= \frac{\|a_2\|}{\|w_2\|} w_2 \end{aligned}$$

Proof:

$\rightarrow$ Follows by direct substitution.

$\leftarrow$ Let $q_1 = \frac{a_1^* w_1}{|a_1^* w_1|}$, since this is well defined. Likewise, choose $d_2 = \dfrac{\|w_2\|}{\|a_2\|}$. By assumption, d_2 is well defined, and nonzero.

Obviously, $\|w_2\| = \|z_2\|$, so let Q_2 be the rotation that takes w_2 into z_2. A quick calculation shows that Q_2 also rotates b_2 into a_2.

$$Q_2 b_2 = \frac{1}{d_2} Q_2 w_2 = \frac{1}{d_2} z_2 = a_2$$

Next, we calculate $a_1^* z_1$. Plugging in gives $a_1^* z_1 = |a_1^* w_1|$. By assumption, this is nonzero, hence Lemma 13.2 yields a hermitian, positive definite D_1 such that $D_1 a_1 = z_1$. As we hope, D_1 takes b_1 into w_1 too.

$$D_1 b_1 = \bar{q}_1 D_1 a_1 = \bar{q}_1 z_1 = w_1$$

Defining D and Q in the obvious manner completes the proof. ♯

This gives us the main theorem.

Theorem 13.9 *Let $M \in \mathbf{C}^{n\times n}$ be given, and let $\boldsymbol{\Delta}$ be the two block ($s = 1, f = 1$) structure defined above, with block sizes r_1 and m_1, where $r_1 + m_1 = n$. Suppose $\beta > 0$ is given. Then there exists $Q \in \mathcal{Q}$, $D \in \mathcal{D}$, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbf{C}^n$, $X_\perp \in \mathbf{C}^{n\times(n-1)}$ such that*

$$\begin{aligned} \|x\| = 1, x_1 \neq 0\ ,\ x_2 \neq 0 \\ x^* X_\perp = 0\ ,\ X_\perp x = 0 \\ QDMD^{-1} = \beta x x^* + X_\perp \end{aligned} \tag{13.25}$$

if and only if there exists nonzero vectors $z_1, w_1, b_1, a_1 \in \mathbf{C}^{r_1}$ and $z_2, w_2, b_2, a_2 \in \mathbf{C}^{m_1}$ with $a_1^* w_1 \neq 0$ and

$$\begin{aligned} \beta a &= Mb \\ z_1 &= \frac{a_1^* w_1}{|a_1^* w_1|} w_1 \\ z_2 &= \frac{\|w_2\|}{\|a_2\|} a_2 \\ \beta w &= M^* z \\ b_1 &= \frac{w_1^* a_1}{|w_1^* a_1|} a_1 \\ b_2 &= \frac{\|a_2\|}{\|w_2\|} w_2. \end{aligned} \tag{13.26}$$

Remark: In order to find decompositions using the representation this theorem allows (equation (13.26) - free of Q's and D's), we **can** restrict ourselves to unit vectors a, b, z, w. Why? Suppose we find nonzero vectors satifying (13.26). Examining these equations, it is clear that scaling z and w by $\alpha \neq 0$ and scaling b and a by $\beta \neq 0$ does not affect any of the equalities in (13.26). Since these equations **always** imply that $\|z\| = \|w\|$, and $\|a\| = \|b\|$, we can indeed look only at unit vectors.

In the above theorem, we have written the conditions (13.26) in a suggestive manner. We will attempt to find solutions to (13.26) in a iterative fashion. In particular, for $i = 1, 2,$

let vectors $a_{i_k}, b_{i_k}, z_{i_k}$, and w_{i_k} evolve as

$$\begin{aligned}
\tilde{\beta}_{k+1} a_{k+1} &= M b_k \\
z_{1_{k+1}} &= \frac{w_{1_k}^* a_{1_{k+1}}}{|w_{1_k}^* a_{1_{k+1}}|} w_{1_k} \\
z_{2_{k+1}} &= \frac{\|w_{2_k}\|}{\|a_{2_{k+1}}\|} a_{2_{k+1}} \\
\hat{\beta}_{k+1} w_{k+1} &= M^* z_{k+1} \\
b_{1_{k+1}} &= \frac{a_{1_{k+1}}{}^* w_{1_{k+1}}}{|a_{1_{k+1}}^* w_{1_{k+1}}|} a_{1_{k+1}} \\
b_{2_{k+1}} &= \frac{\|a_{2_{k+1}}\|}{\|w_{2_{k+1}}\|} w_{2_{k+1}}
\end{aligned} \tag{13.27}$$

where $\tilde{\beta}_{k+1}$ and $\hat{\beta}_{k+1}$ are chosen > 0, so that $\|a_{k+1}\| = \|w_{k+1}\| = 1$.

Note also that if the initial b and w vectors that start the iteration are unit vectors, then at every step, all vectors, a, b, z, and w will be unit length.

13.a There are many other iterative algorithms besides (13.27) that have decompositions (Theorem 13.7) as equilibrium points. For instance, simply rearranging the order of our iteration in (13.27) will yield a different algorithm, yet decompositions are still the equilibrium points. What we really want is an algorithm where the only stable equilibrium points are decompositions with large (relative to $\mu(M)$) converged β values. Other iterations schemes may be better suited toward this goal - discovering them will give a better lower bound algorithm.

13.b Potential problems are:

- $Mb_k = 0$ ($M^* z_k = 0$), then a_{k+1} (w_{k+1}) is not well defined.
- $a_{1_k}^* w_{1_k} = 0$, then the vectors $z_{1_{k+1}}$ and/or $b_{1_{k+1}}$ are not well defined.
- Either $\|w_{2_k}\| = 0$ or $\|a_{2_k}\| = 0$, making b_{2_k} and/or z_{2_k} not well defined.

The heuristic fix when any of these happen is to restart the algorithm at a different initial condition.

13.c If everything goes ok, and all of the indexed quantities converge, then we must have $\tilde{\beta} = \hat{\beta}$. This is easy to see. Suppose the equations in (13.26) are satisfied (convergence of the algorithm in (13.27)), but the β associated with b and a is $\tilde{\beta}$ and the β associated with z and w is $\hat{\beta}$. The converged equations imply that there exists a $Q \in \mathcal{Q}$ and $D \in \mathcal{D}$ such that $QDMD^{-1}(Db) = \tilde{\beta}(Db)$ and $(QDMD^{-1})^*(Db) = \hat{\beta}(Db)$. Since the β's are real, they must be equal.

13.d

- If there is only the first block, which is a scalar times identity block, the iteration would be a power iteration for the largest (in magnitude) eigenvalue of the matrix M. Since μ for 1 scalar times identity block is the spectral radius, the algorithm we have proposed reduces to a valid algorithm in the special case of 1 scalar times identity block.
- If there is only the second block, which is a full block, the iteration becomes a eigenvalue power algorithm for M^*M, hence it will give the largest singular value of M. Again, with respect to this specific block structure, this is what we want.

Hence, the iteration we have proposed is a even mix of two separate, well understood iterations. Both of these converge to the largest eigenvalue/singular value. Therefore, we are led to guess (incorrectly) that this algorithm will converge to the largest β for which a decomposition described in Theorem 13.4 exists.

Extensive computational experience has led to the following conclusions:

1. The difficulties described in 13.b above do not *seem* to occur in practice. While it is easy to construct matrices where these problems happen, running the algorithm on frequency responses of actual closed loop systems has not been a problem.
2. Limit cycles occur more often when there are large scalar times identity blocks. The presence of a stable limit cycle does not immediately give rise to a lower bound for μ.
3. If $s = 0$ (and often times when $s > 0$), the algorithm usually performs well, converging quickly, and providing a lower bound which is better than $\rho(M)$. We have successfully run tests on 40×40 complex matrices with up to 40 complex uncertainties.
4. **The promising properties described above are not always true.** We have examples of a stable equilibrium point with the corresponding $\beta < \rho(M)$. With lack of any further insight, we do not bother to reproduce this here. The block structure was five 1×1 blocks.
5. In general, there are several stable equilibrium points, **with different values of** σ. This is to be contrasted with the conventional power algorithms for ρ and $\bar{\sigma}$, where only the largest ones are stable.

13.5 Choosing starting vectors

This section heuristically addresses the question of "what should the starting vectors be?" To motivate what follows, suppose that $\mu(M) = \inf_{D \in \mathcal{D}_e} \bar{\sigma}\left(DMD^{-1}\right)$, and that the infimum is achieved by D_o. Then, from Theorem 13.4, we must have $0 \in \nabla_{D_o M D_o^{-1}, \bar{\sigma}}$. Therefore, if

$$\tilde{M} := D_o M D_o^{-1} = \mu U V^* + U_2 \Sigma_2 V_2^* \tag{13.28}$$

is a singular value decomposition, there is a $\eta \in \mathbf{C}^r$ and $Q \in \mathcal{Q}$ such that

$$\begin{aligned} \tilde{M} V \eta &= \mu Q V \eta \\ \tilde{M}^* (Q V \eta) &= \mu V \eta \end{aligned} \tag{13.29}$$

Hence, with respect to $\tilde{M}$, (which has $\mu(\tilde{M}) = \mu(M)$), the vectors

$$\begin{aligned} b &:= V \eta \\ w &:= V \eta \end{aligned} \tag{13.30}$$

are the correct vectors for the decompostion. We therefore propose the following.

1. Using a cheap method, [Osb], find a D_{so} that nearly minimizes $\inf_{D \in \mathcal{D}_e} \bar{\sigma}\left(DMD^{-1}\right)$
2. Absorb this into M, ie., define $\tilde{M} := D_{so} M D_{so}^{-1}$
3. choose $b_1 = w_1$ to be a right singular vector associated with $\bar{\sigma}\left(\tilde{M}\right)$
4. perform the iteration on $\tilde{M}$ with these starting vectors

14 Example

To conclude, we analyze the robustness of a nominally stable system subject to structured perturbations. The example has no physical interpretation, and is intended only to illustrate the various robustness theorems we have presented.

The system G (which can be interpreted as $F_l(P(s), K(s))$ as in the previous section) is given below. It has 4 states, and 9 inputs and 9 outputs.

```
Matrix : G.a
states 4
            x1          x2          x3          x4
  x1 -6.405e-01 -5.471e+00 -4.185e+00  2.198e+00
  x2  0.000e+00 -3.000e+00  0.000e+00  0.000e+00
  x3  2.198e+00 -5.098e+00 -2.627e+00  4.185e+00
  x4 -1.987e+00 -6.756e+00 -6.384e+00  1.558e+00

Matrix : G.b
states 4   inputs 9
            u1          u2          u3          u4          u5          u6          u7          u8          u9
  x1  5.336e-01  0.000e+00  0.000e+00  0.000e+00  0.000e+00  0.000e+00  0.000e+00  0.000e+00  7.000e-01
  x2  0.000e+00  0.000e+00  0.000e+00  0.000e+00  2.668e-01  8.893e-02  0.000e+00  0.000e+00  7.000e-01
  x3  5.336e-01  0.000e+00  5.336e-01  2.668e-01  0.000e+00  0.000e+00  0.000e+00  0.000e+00  7.000e-01
  x4  0.000e+00  5.336e-01  0.000e+00  0.000e+00  0.000e+00  0.000e+00 -8.893e-02  8.893e-02  7.000e-01

Matrix : G.c
states 4   outputs 9
            x1          x2          x3          x4
  y1  2.500e-01  0.000e+00  2.500e-01  0.000e+00
  y2  0.000e+00  2.500e-01  0.000e+00  0.000e+00
  y3  0.000e+00  0.000e+00  0.000e+00  2.500e-01
  y4  0.000e+00  5.000e-01  0.000e+00  0.000e+00
  y5  0.000e+00  5.000e-01  0.000e+00  5.000e-01
  y6  0.000e+00  0.000e+00  0.000e+00  1.500e+00
  y7  0.000e+00  1.500e+00  0.000e+00  0.000e+00
  y8  0.000e+00  0.000e+00  1.500e+00  1.500e+00
  y9  0.000e+00 -1.000e+00  0.000e+00  1.000e+00
```

The first 8 inputs and outputs are associated with the perturbation structure, $\mathbf{\Delta}$, $s = 3, f = 0, r_1 = 3, r_2 = 2, r_3 = 3$. The last input and output correspond to the exogenous disturbance and resulting error. Hence, for **robust performance** calculations, we will append a 1×1 full block to $\mathbf{\Delta}$ for the performance calculation. For notational purposes, we partition $G(s)$ into

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \tag{14.1}$$

where $G_{11}(s)$ is 8×8, and $G_{22}(s)$ is 1×1.

We first calculate the robustness with respect to linear, time invariant perturbations, using the frequency domain techniques described in the section 11. This is done via a μ test on $G_{11}(j\omega)$. At each frequency point, we calculate the $\bar{\sigma}(DMD^{-1})$ upper bound, and a lower bound using the algorithm described in section 13.

- Figure 14.1 is simply the singular values of $G_{11}(j\omega)$ versus ω. This implies that for any **unstructured** (any 8×8), stable perturbation, with induced norm from $L_2 \rightarrow L_2$ less than $\frac{1}{3.92}$, the perturbed closed loop system is stable. The Nyquist argument also shows that **there is** a linear, time invariant, unstructured, stable perturbation, Δ_u with $\sup_{\omega \in \mathbf{R}} \bar{\sigma}(\Delta(j\omega)) = \frac{1}{3.92}$ that does cause instability.
- Next, we calculate upper and lower bounds for $\mu(G_{11}(j\omega))$, with respect to the block structure $\mathbf{\Delta} := \{\text{diag}[\delta_1 I_3, \delta_2 I_2, \delta_3 I_3] : \delta_i \in \mathbf{C}\}$. The upper bound is based on the generalized gradient material from section 6.1, and the lower bound is the iterative procedure described in section 13. These two curves are nearly equal, and are shown in Figure 14.2. This implies that for FDLTI perturbations $\Delta(s)$ with the correct structure, the stability is preserved as long as $\|\Delta(s)\|_\infty < \frac{1}{0.64}$, and there is a perturbation on that boundary that does cause instability.

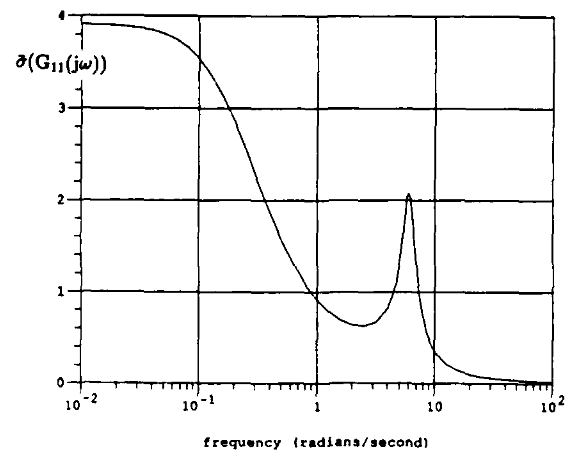


Figure 14.1 Frequency Response Singular Value Plot

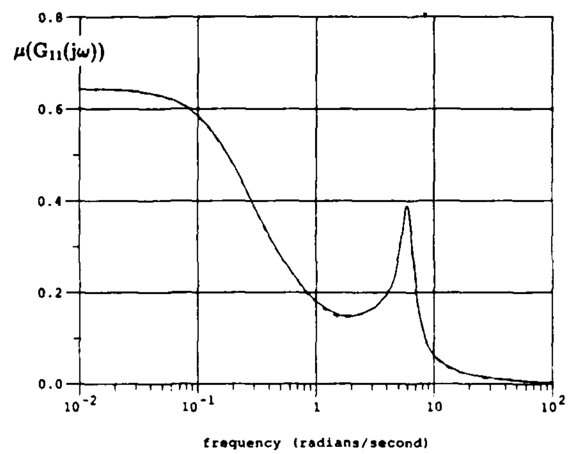


Figure 14.2 Frequency Response μ plot

- What about performance? Nominally, the transfer function G_{22} describes the performance, and this is shown in Figure 14.3. It has a peak value of 0.83. Under perturbations this becomes $F_u(G, \Delta)$. To analyze the degradation of performance due to the uncertainty, we use theorem 11.4, and an augmented block structure $\hat{\mathbf{\Delta}}$

$$\hat{\mathbf{\Delta}} := \{\text{diag}[\Delta, \delta_2] : \Delta \in \mathbf{\Delta}, \delta_2 \in \mathbf{C}\} \tag{14.2}$$

A μ plot of $\mu_{\hat{\mathbf{\Delta}}}(G(j\omega))$ is shown in Figure 14.4. Applying a scaled version of theorem

11.4 implies that for any structured $\Delta(s)$ with $\|\Delta(s)\|_\infty < \frac{1}{1.18}$, the perturbed loop remains stable (we already knew that for an even greater radius from the μ test on G_{11}), **and**, the $\|\cdot\|_\infty$ norm of $F_u(G, \Delta)$ is guaranteed to be ≤ 1.18.

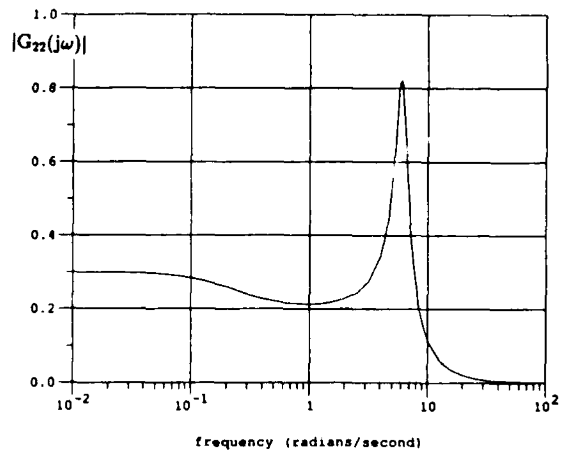


Figure 14.3 Nominal Disturbance to Error Frequency Response

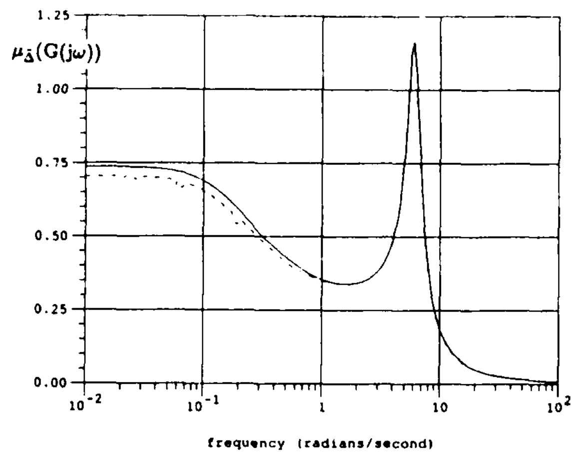


Figure 14.4 μ Plot for Robust Performance

Finally, we consider robust stability to time varying perturbations, using the optimal constant D scaling result from section 10 to minimize the conservatism of the small gain theorem, by taking into account the structure. This will give a sufficient condition for robust stability to time varying, and also cone bounded, nonlinear perturbations as well. (The correct formulas for continuous time systems are given in the appendix, and are in the same spirit as (10.9) and (10.10)). Everything pertains to G_{11}, since we are only concerned with stability. From Figure 14.2, we know that the optimal value satisfies

$$\inf_{D \in \mathcal{D}} \|DG_{11}(s)D^{-1}\|_\infty \geq \sup_{\omega} \inf_{D_\omega \in \mathcal{D}} \bar{\sigma}\left(D_\omega G_{11}(j\omega) D_\omega^{-1}\right) = 0.64 \qquad (14.3)$$

We performed a 1 dimensional search to find the correct value of γ (in equation (10.10)). Our rather crude gradient algorithm indicates that $\gamma \in (0.68, 0.685)$. This is quite close to the frequency varying optimal. The constant D scaling we get from setting $\alpha = 0.685$ is given below.

```
Matrix : D.opt    BLOCK DIAGONAL    rows 8   columns 8

          1           2           3
  1   1.113e+02  -7.359e+01   6.100e+00
  2  -2.662e+01  -1.638e+01   5.527e+01
  3   6.537e+00   7.397e+01   3.210e+01

          4           5
  4   3.716e+01  -3.139e+01
  5   4.586e+01   1.975e+01
```

	6	7	8
6	3.118e+00	-1.112e+01	1.115e+01
7	6.516e+00	-2.682e+00	-3.690e+00
8	1.623e+01	9.812e+00	-8.816e+00

If we scale $G_{11}(j\omega)$ with this constant scaling Figure 14.5 results. Note that the problem is of the sort

$$\inf_{D \in \mathcal{D}} \sup_{\substack{\omega \\ \text{all } \sigma_i}} \sigma_i \left[\text{func}(D, \omega) \right]. \tag{14.4}$$

We expect coalesced behavior at the minimum, and this is exactly what we have. In this instance, though, the coalesced behavior is with respect to the ω variable - Figure 14.5 shows this very clearly.

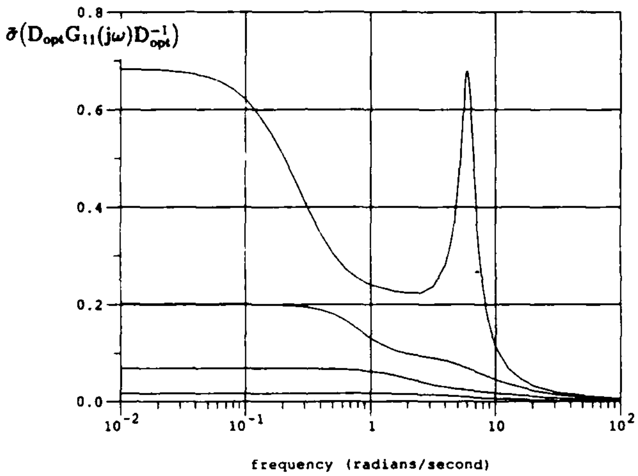


Figure 14.5 Singular Value Plot with Optimal, Constant D Scales

As we noted, this example has no physical significance, it merely demonstrates several of the different ideas we have covered in this report, namely frequency domain, and state space μ techniques, as well as the optimal constant scaling material of section 10 for time varying perturbations. Several realistic examples using μ have appeared in the literature, including [DoyLP] and [DoySE]. The emphasis in each of these is a particular example - the various uses and interpretations possible with the different μ calculations are not the main issues.

15 Appendix

15.1 Star Products

Recall the example from section 8.3.1 and the results on uncertain difference equations in section 9. Both of these were done in discrete time, since in that domain, the unit disk is important, and disks are what μ is all about. This section shows that the well known bilinear transformation yields results analogous to the above for continuous time systems. We begin with a generalization of the LFT, called "the star product" which is found in [Red].

Suppose Q and M are two complex matrices, which we partition as

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \qquad M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

We are a little cavalier about the dimensions here. We only require that the matrix product $Q_{22}M_{11}$ makes sense and is square. Obviously then, the product $M_{11}Q_{22}$ also makes sense and also is square. If the matrix $I - Q_{22}M_{11}$ is invertible, then we say the *star product* $Q{*}M$, is well defined, and is given by

$$Q{*}M := \begin{bmatrix} F_l\left(Q, M_{11}\right) & Q_{12}\left(I - M_{11}Q_{22}\right)^{-1} M_{12} \\ M_{21}\left(I - Q_{22}M_{11}\right)^{-1} Q_{21} & F_u\left(M, Q_{22}\right) \end{bmatrix}. \qquad (15.1)$$

Note that this definition is dependent on the partitioning of the matrices Q and M above. In fact it may be well defined for one partition and not well defined for another. However, we will not explicitely show this dependence, as it is always clear from the context.

In a block diagram, the star product (15.1) has a natural interpretation: it is simply the matrix relating $\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ to $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ as shown below.

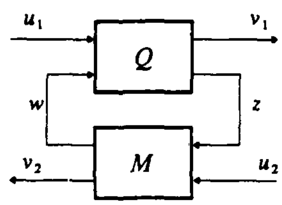


Figure 15.1 Star Product of Two Matrices

The assumption that $I - Q_{22}M_{11}$ is invertible implies that for any vectors u_1 and u_2, there exist unique vectors z and w satisfying the loop equations. When working with star products, *it is much easier to manipulate the diagrams, rather than the equations*, since the diagrams are so intuitive, however a little care must be exercised. Consider the loop below.

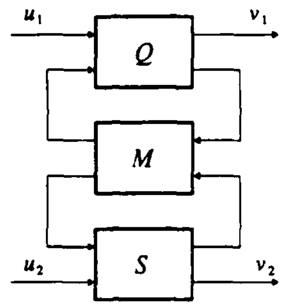


Figure 15.2 Associativity of Star Products

Should this be viewed as $(Q*M)*S$ or $Q*(M*S)$? Well, when looking at it pictorially, it appears to make no difference. But, we have to be careful about the invertibility of the necessary matrices. For example, suppose all the matrices are 2×2, and that $Q_{22} = 0.5$, $M = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, and $S_{11} = 1$ (the rest of Q and S are irrelevent). Certainly $Q*M$ is okay, and since $[Q*M]_{22} = 2$, the quantity $1-[Q*M]_{22}S_{11}$ is invertible, and therefore $(Q*M)*S$ is well defined. But, since the star product $M*S$ is not even defined, we cannot compare the first expression, $(Q*M)*S$ to $Q*(M*S)$.

So, if we want to have associativity (which is what to need to manipulate the diagrams, rather than working via the fairly messy definition), both $Q*M$ and $M*S$ should be well defined. This requires that both $I - Q_{22}M_{11}$ and $I - M_{22}S_{11}$ are invertible. In this case, the next lemma and corollary show that if either $(Q*M)*S$ or if $Q*(M*S)$ is well defined, then they are equal.

Lemma 15.1 *If both* $I - Q_{22}M_{11}$ *and* $I - M_{22}S_{11}$ *are invertible, then the quantity* $I - F_u(M, Q_{22})S_{11}$ *is invertible if and only if* $I - F_l(M, S_{11})Q_{22}$ *is invertible.*

Proof: We manipulate determinants: $\det\left[I - F_u(M, Q_{22})S_{11}\right] \neq 0$

$$
\begin{aligned}
&\leftrightarrow \det\left\{I - \left[M_{22} + M_{21}Q_{22}(I - M_{11}Q_{22})^{-1}M_{12}\right]S_{11}\right\} \neq 0 \\
&\leftrightarrow \det\left\{I - M_{22}S_{11} - M_{21}Q_{22}(I - M_{11}Q_{22})^{-1}M_{12}S_{11}\right\} \neq 0 \\
&\leftrightarrow \det\left\{I - M_{21}Q_{22}(I - M_{11}Q_{22})^{-1}M_{12}S_{11}(I - M_{22}S_{11})^{-1}\right\} \neq 0 \\
&\leftrightarrow \det\left\{I - M_{12}S_{11}(I - M_{22}S_{11})^{-1}M_{21}Q_{22}(I - M_{11}Q_{22})^{-1}\right\} \neq 0 \\
&\leftrightarrow \det\left\{I - M_{11}Q_{22} - M_{12}S_{11}(I - M_{22}S_{11})^{-1}M_{21}Q_{22}\right\} \neq 0 \\
&\leftrightarrow \det\left\{I - F_l(M, S_{11})Q_{22}\right\} \neq 0 \; \sharp
\end{aligned}
$$

This implies the corollary.

Corollary 15.2 *Let Q, M, and S be given. If $Q*M$ and $M*S$ are each well defined (ie. $I - Q_{22}M_{11}$ and $I - M_{22}S_{11}$ are invertible), then $(Q*M)*S$ is well defined if and only if $Q*(M*S)$ is well defined. Furthermore, if they are well defined, then they are equal.*

These star products have many interesting properties discovered in [Red]. We will not pursue them here. In the next section though, we use star products to translate the discrete time results from sections 8.3.1 and 9 to analogous continuous time results.

15.2 Continuous time results

In this section we show that the well known bilinear transformation, along with the star product, yields results for continuous time systems.

Let $n > 0$ be an integer, and define a matrix $\mathbf{B}$ by

$$\mathbf{B} := \begin{bmatrix} I_n & \sqrt{2}\, I_n \\ \sqrt{2}\, I_n & I_n \end{bmatrix}$$

Suppose $A \in \mathbf{C}^{n \times n}$. It is simple to relate the eigenvalues of A and $F_l(\mathbf{B}, A)$. In particular,

Lemma 15.3 *Let $\lambda_i, i = 1, \ldots, n$, denote the eigenvalues of A. Then $\mathrm{Re}(\lambda_i) < 0$ for each i if and only if $I - A$ is invertible, and $\rho[F_l(\mathbf{B}, A)] < 1$.*

Similarly, we have a matrix version of the bilinear transformation.

Lemma 15.4 *Suppose $A \in \mathbf{C}^{n \times n}$. Let $A_H := \frac{1}{2}(A + A^*)$. Then $A_H < 0$ if and only if $I - A$ is invertible and $\bar{\sigma}(F_l(\mathbf{B}, A)) < 1$.*

Proof: Suppose that $I - A$ is invertible and $\bar{\sigma}(F_l(\mathbf{B}, A)) < 1$. Then

$$\begin{aligned} \bar{\sigma}(F_l(\mathbf{B}, A)) < 1 \quad & \text{iff} \quad \bar{\sigma}\left((I + A)(I - A)^{-1}\right) < 1 \\ & \text{iff} \quad (I - A^*)^{-1}(I + A^*)(I + A)(I - A)^{-1} < I \\ & \text{iff} \quad (I + A^*)(I + A) < (I - A^*)(I - A) \\ & \text{iff} \quad A^* + A < -A^* - A \\ & \text{iff} \quad A_H < 0 \end{aligned}$$

Reversing the steps gives the proof for the other direction. $\sharp$

Now, let us apply this to a class of uncertain differential equations, as we did for the discrete case in section 9. To set it up, let $M \in \mathbf{C}^{(n+m)\times(n+m)}$ be given, along with a $m \times m$ block structure $\mathbf{\Delta}$, such that $\mu_{\mathbf{\Delta}}(M_{22}) < 1$. We are interested in solutions $x(t) \in \mathbf{C}^n$ that evolve according to

$$\dot{x} = F_l(M, \Delta(t))\, x$$

where the function $\Delta(\cdot)$ is piecewise continuous, say. We assume that the nominal system is known to be stable, hence all of the eigenvalues of M_{11} have negative real parts. Consider the following three assumptions on $\Delta(\cdot)$

(a.1) For all t, $\Delta(t) \in \mathbf{\Delta}$

(a.2) For all t, $\bar{\sigma}(\Delta(t)) \leq 1$

(a.3) $\Delta(\cdot)$ is constant - it does not vary with t

Now, (a.3) implies that the system is time invariant, so we just need to check that the dynamic matrix, $F_l(M, \Delta)$ is hurwitz for each allowable Δ. Equivalently, via Lemma 15.3, we need to check that $\rho[F_l(\mathbf{B}, F_l(M, \Delta))] < 1$ for all allowable Δ. This is displayed in block diagram form below on the left.

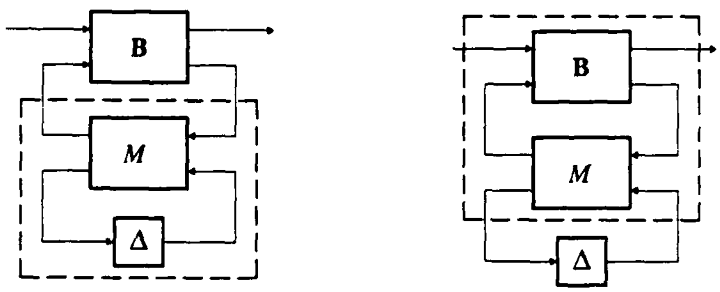


Figure 15.3 Uncertain Differential Equations

We would like to exchange the order, and evaluate whether $\rho[F_l(\mathbf{B}*M), \Delta] < 1$ for all Δ, because this is just a μ test on $\mathbf{B}*M$. This is illustrated above right. Theorem 15.5 handles this.

Theorem 15.5 *Define* $\tilde{\mathbf{\Delta}} := \{\mathrm{diag}[\delta I_n, \Delta] : \delta \in \mathbf{C}, \Delta \in \mathbf{\Delta}\}$. *Then, with the above assumptions, the differential equation* $\dot{x} = F_l(M, \Delta)x$ *is stable for all fixed* $\Delta \in \mathbf{\Delta}$, *with* $\bar{\sigma}(\Delta) \leq 1$ *if and only if* $\mu_{\tilde{\mathbf{\Delta}}}(\mathbf{B}*M) < 1$.

Proof: Since the nominal matrix M_{11} has all of its eigenvalues in the left half plane, the star product $\mathbf{B}*M$ is well defined. Also by assumption, $\mu_{\mathbf{\Delta}}(M_{22}) < 1$, hence for every

$\Delta \in \mathbf{\Delta}$, with $\bar{\sigma}(\Delta) \leq 1$, the LFT $F_l(M, \Delta)$ is well defined. Hence, the standing assumptions of lemma 15.1 (or corollary 15.2) are satisfied.

$\leftarrow$ By hypothesis, if $\Delta \in \mathbf{\Delta}$, and $\bar{\sigma}(\Delta) \leq 1$, then $I - [\mathbf{B}*M]_{22}\Delta$ is invertible, and hence so is $I - F_l(M, \Delta)$. Therefore, by corollary 15.2, for all such Δ, we have

$$F_l(\mathbf{B}, F_l(M, \Delta)) = F_l(\mathbf{B}*M, \Delta)$$

Therefore

$$\max_{\substack{\Delta \in \mathbf{\Delta} \\ \bar{\sigma}(\Delta) \leq 1}} \rho[F_l(\mathbf{B}, F_l(M, \Delta))] = \max_{\substack{\Delta \in \mathbf{\Delta} \\ \bar{\sigma}(\Delta) \leq 1}} \rho[F_l(\mathbf{B}*M, \Delta)] < 1$$

where the last inequality comes from the assumption that $\mu_{\tilde{\Delta}}(\mathbf{B}*M) < 1$, and the robust performance theorem, 5.2, applied to $\mathbf{B}*M$, with the block structure $\tilde{\mathbf{\Delta}}$. Hence, using lemma 15.1 shows that the eigenvalues of $F_l(M, \Delta)$ are in the left half plane for each Δ.

$\rightarrow$ Same type of argument. $\sharp$

Similar results are obtained for the other situations. We collect them here.

Lemma 15.6 *There is a single Lyapunov matrix for the entire set of "A" matrices*

$$\{F_l(M, \Delta) : \Delta \in \mathbf{\Delta}, \bar{\sigma}(\Delta) \leq 1\}$$

if and only if

$$\inf_{\substack{T \in \mathbf{C}^{n \times n} \\ T \text{invertible}}} \mu_{\hat{\Delta}}\left(\begin{bmatrix} T & 0 \\ 0 & I \end{bmatrix} (\mathbf{B}*M) \begin{bmatrix} T^{-1} & 0 \\ 0 & I \end{bmatrix}\right) < 1$$

where $\hat{\mathbf{\Delta}} := \{\text{diag}[\Delta_1, \Delta] : \Delta_1 \in \mathbf{C}^{n \times n}, \Delta \in \mathbf{\Delta}\}$.

Lemma 15.7 *Let* $\tilde{\mathcal{D}} := \{\text{diag}[D_1 I_n, D] : D_1 \in \mathbf{C}^{n \times n}, \text{ invertible}, D \in \mathcal{D}\}$, *where* $\mathcal{D}$ *is the appropriate scaling set for the block structure* $\mathbf{\Delta}$. *Then a sufficient condition for Lemma 15.6 is*

$$\inf_{\tilde{D} \in \tilde{\mathcal{D}}} \bar{\sigma}\left(\tilde{D}\, \mathbf{B}*M\, \tilde{D}^{-1}\right) < 1$$

We can also use $\mathbf{B}$ and the star product to switch between z and s domains for transform results.

Lemma 15.8 *Let A, B, C, D be a state space realization of a stable, continuous time transfer function $G(s)$, with m inputs and m outputs. We assume that the matrix A is Hurwitz. Define the matrix M as*

$$M := \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

Let $\mathbf{\Delta} := \{\mathrm{diag}\,[\delta_1 I_n, \Delta_2] : \delta_1 \in \mathbf{C}, \Delta \in \mathbf{C}^{m\times m}\}$. Then

$$\|G\|_\infty < 1 \text{ iff } \mu_\Delta (\mathbf{B}{*}M) < 1$$

Lemma 15.9 *Let $G(s) := D + C\,(sI - A)^{-1}\,B$ be a m input, m output, rational, stable transfer function. Suppose $\mathbf{\Delta}_2$ is a $m \times m$ block structure as in (3.1), and let $\mathcal{D}_2$ be the corresponding scaling set. For $\alpha > 0$, define*

$$M^\alpha := \begin{bmatrix} A & B \\ \alpha C & \alpha D \end{bmatrix} \tag{15.2}$$

Define $\gamma \in \mathbf{R}$ by

$$\gamma := \sup_{\alpha>0} \left\{ \alpha : \inf_{\substack{D_1 \text{ invertible} \\ D_2 \in \mathcal{D}_2}} \bar{\sigma} \left(\begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} (\mathbf{B}{*}M^\alpha) \begin{bmatrix} D_1^{-1} & 0 \\ 0 & D_2^{-1} \end{bmatrix} \right) < 1 \right\}. \tag{15.3}$$

Then

$$\inf_{D_2 \in \mathcal{D}_2} \sup_{\substack{s \in \mathbf{C} \\ \mathrm{Re}(s) \geq 0}} \bar{\sigma}\left(D_2 G(s) D_2^{-1}\right) = \frac{1}{\gamma}\sharp \tag{15.4}$$

15.3 Convexity Lemma

The following lemma gives a sufficient condition for a continuous function from $\mathbf{R} \to \mathbf{R}$ to be convex. It is fairly intuitive, and comes from [ChuD]

Lemma 5.2 *Let $f : \mathbf{R} \to \mathbf{R}$ be a continuous function, and suppose for each $t_o \in \mathbf{R}$, there exists function $g_{t_o} \in C^2$ (continuously twice differentiable), $g_{t_o} : \mathbf{R} \to \mathbf{R}$, such that $f(t_o) = g_{t_o}(t_o)$, $f(t) \geq g_{t_o}(t)$ for all $t \in \mathbf{R}$ and $\frac{d^2 g_{t_o}}{dt^2}\Big|_{t=t_o} \geq 0$. Then f is a convex function.*

Proof: Suppose f is not convex. Then there exist $x, y \in \mathbf{R}$, $x < y$, and $\lambda \in (0,1)$ such that

$$f\left((1-\lambda)x + \lambda y\right) > (1-\lambda) f(x) + \lambda f(y)$$

Let β be the largest difference this assumes, ie.

$$\beta = \max_{\alpha \in [0,1]} \left[f\left((1-\alpha)x + \alpha y\right) - (1-\alpha) f(x) - \alpha f(y) \right]$$

and let $\bar{\lambda}$ be the largest value in $[0,1]$ that achieves β. Obviously, since $\beta > 0$, $\bar{\lambda} \in (0,1)$. Define $\bar{w} := (1-\bar{\lambda})x + \bar{\lambda}y$. Hence f is continuous, satisfies $f(\bar{w}) = \beta$, and lies in the shaded region as shown below (shaded region includes its boundary for $t < \bar{w}$, and does not include its boundary for $t > \bar{w}$).

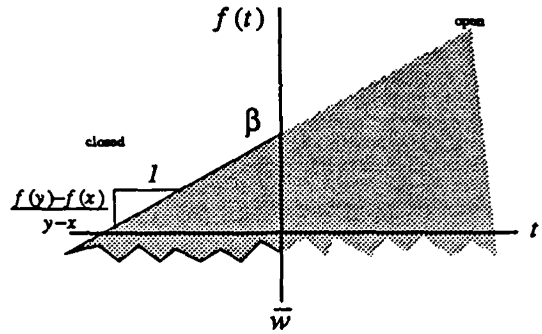


Figure 15.4 Diagram for Lemma 5.2

Now, let g be any function in C^2 with $g(\bar{w}) = \beta$, and $\left.\dfrac{d^2 g}{dt^2}\right|_{t=\bar{w}} \geq 0$. Obviously, there are points w arbitrarily close to $\bar{w}$ such that $f(w) < g(w)$. So, by contrapositive, we have proven the lemma. ♯

16 References

Bar: B.R. Barmish, "Invariance of the Strict Hurwitz property for polynomials with perturbed coefficients," *IEEE Trans. Auto. Control*, vol. AC-29, pp. 935-936, 1984.

BarHL: A.C. Barlett, C.V. Hollot, and H. Lin, "Root locations of an Entire Polytope of Polynomials: It suffices to check the edges," *Mathematics of Control, Signals, and Systems*, vol. 1, no. 1, 1988.

BoyD: S. Boyd and C.A. Desoer, "Subharmonic Functions and Performance Bounds on Linear Time Invariant Feedback Systems," *UC Berkeley/ERL Memo M84/51*, 1984.

CheD M.J. Chen and C.A. Desoer, "Necessary and sufficient condition for robust stability of linear distributed feedback systems," *Int. J. Control*, vol. 35, no. 2, pp. 255-267, 1982.

ChuD: C. Chu and J. Doyle, private communication, 1985.

DanKL: R.W. Daniel, B. Kouvaritakis, and H. Latchman, "Principal direction alignment: a geometric framework for the complete solution to the μ problem," *IEE Proceedings*, vol. 133, Part D, No. 2, pp. 45-56, March 1986.

deGS: R.R.E. de Gaston and M.G. Safonov, "Calculation of the Multiloop Stability Margin," *Amer. Control Conf.*, June 1986, Seattle, pp. 761-770.

DesV: C.A. Desoer and M. Vidyasagar, "Feedback Systems: Input-Output Properties," *Academic Press*, New York, 1975.

Doy: J.C. Doyle, "Analysis of feedback systems with structured uncertainties", *IEE Proceedings*, vol. 129, Part D, No. 6, pp. 242-250, Nov. 1982.

Doy2: J.C. Doyle, "Structured uncertainty in control system design", IEEE CDC, December 1985, Ft. Lauderdale.

DoyLP: J. Doyle, K. Lenz, and A. Packard, "Design Examples using μ synthesis: Space Shuttle Lateral Axis FCS during reentry," *IEEE CDC*, December 1986, pp. 2218-2223.

DoyP: J.C. Doyle and A. Packard, "Uncertain Multivariable Systems from a State Space Perspective", *American Control Conf.*, June 1987, pp. 2147-2152.

DoyS: J.C. Doyle and G. Stein, "Multivariable feedback design: Concepts for a classical/modern synthesis," *IEEE Trans. Auto. Control*, vol. AC-26, pp. 4-16, Feb. 1981.

DoySE: J.C. Doyle, R.S. Smith, and D.F. Enns, "Control of Plants with Input Saturation Nonlinearities," *American Control Conf.*, June 1987, pp 1034-1039.

DoyWS: J.C. Doyle, J. Wall and G. Stein, "Performance and robustness analysis for structured uncertainty," in *Proc. 21st IEEE Conf. Decision Contr.*, Dec. 1982, pp. 629-636.

DunS: N. Dunford and J.T. Schwartz, "Linear Operators," Interscience Publishers, Wiley, New York, 1958 and 1963.

FanT1: M.K.H. Fan and A.L. Tits, "Characterization and Efficient Computation of the Structured Singular Value", *IEEE Trans. Auto. Control*, vol. AC-31, no. 8, pp. 734-743, Aug. 1986.

FanT2: M.K.H. Fan and A.L. Tits, "m-Form Numerical Range and the Computation of the Structured Singular Value," *IEEE Trans. Auto. Control*, vol. AC-33, no. 3, pp. 284-289, March 1988.

FanT3: M.K.H. Fan and A.L. Tits, "Toward a Structured Singular Value Decomposition", in *Proc. 26'th IEEE Conf. Decision and Contr.*, Dec. 1987, pp. 1742-3.

FanTD: M.K.H. Fan, A.L. Tits, and J. Doyle, *Amer. Control Conf.*, Atlanta, June 1988.

FooP: Y.K. Foo and I. Postlethwaite, "Extensions of the small-μ test for robust stability," *IEEE Trans. Auto. Control*, vol. 33, no. 2, Feb. 1988, pp. 172-176.

FreLC J.S. Freudenberg, D.P. Looze and J.B. Cruz,, "Robustness analysis using singular value sensitivities", *Int. J. Control*, vol. 35, no. 1, pp. 95-116, 1982.

Gil: E.G. Gilbert, "An Iterative Method for Computing the Minimum of a Quadratic Form on a Convex set", *SIAM Journal on Control* , 4, 1966, pp. 61-80.

Hau: J.E. Hauser, "Proximity Algorithms: Theory and Implementation," *UC Berkeley/ERL Memo,* M86/53, May, 1986.

Hel: W. Helton, "A numerical method for computing the structured singular value," *Systems and Control Letters*, pp. 21-26, 1988.

Kat: T. Kato, "A Short Introduction to Perturbation Theory for Linear Operators," 1982, Springer Verlag, New York.

Kha: V.L. Kharitonov, "Asymptotic Stability of an Equilibrium Position of a Family of Systems of Linear Differential Equations," *Differential Equations*, vol. 14, 1979, pp. 1483-1485.

MinAD: R.J. Minnichelli, J.J. Anagnost and C.A. Desoer, "An Elementary Proof of Kharitonov's Stability Theorem with Extensions," *UC Berkeley/ERL Memo M87/78*, Nov. 1987.

MorD: B. Morton and J. Doyle, private communication, 1985.

MorM: B. Morton and R. McAfoos, "A Mu-Test for Real-Parameter Variations," *American Control Conf.*, Boston, 1985.

NetU: C.N. Nett and J.A. Uthgenannt, "An Explicit Formula and an Optimal Weight for the 2 Block Structured Singular Value Interaction Measure," *Amer. Control Conf.*, June 1987, Minneapolis, pp. 506-511.

NobD: B. Noble and J.W. Daniel, "Applied Linear Algebra", 1977, 2nd edition, Prentice-Hall, Englewood Cliffs, New Jersey.

Osb: E.E. Osborne, "On Preconditioning of Matrices," *J. Assoc. Comp. Mach.*, 7, pp. 338-345, 1960.

Red: R. Redheffer, "Inequalities for a matrix Riccati equation," *Journal of Mathematics and Mechanics*, vol. 8, no. 3, 1959.

Roc: R.T. Rockafellar, "Convex Analysis," 1970, Princeton University Press, Princeton, New Jersey.

Rud: W. Rudin, "Real and Complex Analysis," McGraw Hill Book Company, New York, 1966.

Sae: M. Saeki, "A method of Robust Stability Analysis with Highly Structured Uncertainties," IEEE trans. Auto. Control, vol. AC-31, no. 10, pp. 935-940, 1986.

Saf1: M.G. Safonov, "Stability and Robustness of Multivariable Feedback Systems," *MIT Press*, 1980.

Saf2: M.G. Safonov, "Stability of interconnected systems having slope-bounded nonlinearities," 6th Int. Conf. on Analysis and Optimization of Systems, Nice, France. 1984

SafD: M.G. Safonov and J.C. Doyle, "Minimizing conservativeness of robustness singular values", in *Multivariable Control*, S.G. Tzafestas, Ed. New York: Reidel, 1984.

SidG: A. Sideris and R.R.E. de Gaston, "Multivariable Stability Margin Calculation with Uncertain Correlated Parameters," *IEEE CDC Conf.*, December 1986, Greece, pp. 766-773.

SidP: A. Sideris and R. Sanchez Peña, "Robustness margin calculations with dynamic and real parametric uncertainty," 1988 ACC.

Vid: M. Vidyasagar, "Control System Synthesis: A Factorization Approach," *MIT Press*, Cambridge, MA. 1985.

Wil: J.C. Willems, "Least Squares Stationary Optimal Control and the Algebraic Riccati Equation", *IEEE Trans. Auto. Control*, vol. AC-16, no. 6, pp. 621-634, Dec. 1971.

Wol: P. Wolfe, "Finding the Nearest Point in a Polytope", *Mathematical Programming*, 11, 1976, pp. 128-149.

YedL: R.K. Yedavalli and Z. Liang, "Reduced conservatism in stability robustness bounds by state transformation," *IEEE Trans. Auto. Control*, vol AC-31, no. 12, Sept 1986, pp. 863-866.

Zam: G. Zames, "On the input-output stability of nonlinear time-varying feedback systems, parts I and II," *IEEE Trans. Auto. Control* , vol. AC-11, pp. 228 and 465.

ZhoK: K. Zhou and P. Khargonekar, "Stability robustness bounds for linear state-space models with structured uncertainty," *IEEE Trans. Auto. Control*, vol. AC-31, July 1987, pp. 621-623.

END
DATE
FILMED
8-88
DTIC